GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:08:45 ; Search time 45.463 Seconds
(without alignments)
2075.771 Million cell updates/sec

Title: US-10-650-585-4
Perfect score: 1771
Sequence: 1 MKKKKLEHHHHHHTSAGITK.....TTMRTSSAWRHPQFGGKKKK 334

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1771	100.0	334	5	ABG32182	Abg32182 HCV prote
2	1660	93.7	409	5	ABG32181	Abg32181 HCV prote
3	1589	89.7	303	5	ABG32183	Abg32183 HCV prote
4	1589	89.7	341	5	ABG32187	Abg32187 HCV prote
5	1589	89.7	352	5	ABG32186	Abg32186 HCV prote
6	1589	89.7	380	5	ABG32185	Abg32185 HCV prote
7	1589	89.7	393	5	ABG32184	Abg32184 HCV prote
8	1580	89.2	303	5	ABG32191	Abg32191 HCV prote
9	1579	89.2	303	5	ABG32189	Abg32189 HCV prote
10	1570	88.7	301	5	ABG32190	Abg32190 HCV prote
11	1532	86.5	292	5	ABG32188	Abg32188 HCV prote
12	1531	86.4	2201	5	ABG30601	Abg30601 Hepatitis
13	1531	86.4	2201	5	ABG30591	Abg30591 Hepatitis
14	1531	86.4	2201	5	ABG30600	Abg30600 Hepatitis
15	1531	86.4	2201	5	ABG30581	Abg30581 Hepatitis
16	1531	86.4	2201	5	ABG30593	Abg30593 Hepatitis
17	1531	86.4	2201	5	ABG30582	Abg30582 Hepatitis
18	1531	86.4	2201	5	ABG30580	Abg30580 Hepatitis
19	1531	86.4	2201	5	ABG30587	Abg30587 Hepatitis
20	1531	86.4	2201	5	ABG30599	Abg30599 Hepatitis
21	1531	86.4	2201	5	ABG30594	Abg30594 Hepatitis
22	1531	86.4	2201	5	ABG30598	Abg30598 Hepatitis
23	1531	86.4	2201	5	ABG30595	Abg30595 Hepatitis
24	1531	86.4	3010	5	ABG32458	Abg32458 Hepatitis
25	1531	86.4	3010	5	ABG32459	Abg32459 Hepatitis
26	1531	86.4	3010	5	ABG32451	Abg32451 Hepatitis
27	1531	86.4	3010	5	ABG32455	Abg32455 Hepatitis
28	1531	86.4	3010	5	ABG32457	Abg32457 Hepatitis
29	1531	86.4	3010	5	ABG32460	Abg32460 Hepatitis
30	1531	86.4	3010	5	ABG32461	Abg32461 Hepatitis
31	1531	86.4	3010	5	ABG32454	Abg32454 Hepatitis
32	1531	86.4	3010	5	AAE20477	Aae20477 HCV-S1 fu
33	1531	86.4	3011	5	ABG32456	Abg32456 Hepatitis
34	1528	86.3	2201	5	ABG30586	Abg30586 Hepatitis
35	1528	86.3	2201	5	ABG30589	Abg30589 Hepatitis
36	1528	86.3	2201	5	ABG30583	Abg30583 Hepatitis
37	1528	86.3	2201	5	ABG30588	Abg30588 Hepatitis
38	1528	86.3	2307	3	AAY70064	Aay70064 Recombina
39	1528	86.3	3010	2	AAR68622	Aar68622 HCV prote
40	1528	86.3	3010	2	AAR82694	Aar82694 Partial H
41	1527	86.2	2201	5	ABG30590	Abg30590 Hepatitis
42	1525	86.1	2307	3	AAY70065	Aay70065 Recombina
43	1525	86.1	3010	5	ABG32452	Abg32452 Hepatitis
44	1524	86.1	2201	5	ABG30584	Abg30584 Hepatitis
45	1524	86.1	2201	5	ABG30602	Abg30602 Hepatitis

ALIGNMENTS

RESULT 1
ABG32182
ID ABG32182 standard; protein; 334 AA.
XX
AC ABG32182;
XX
DT 05-NOV-2002 (first entry)
XX
DE HCV protease NS2/3 truncation 4K-6H (904-1206)st-4K.
XX
KW HCV; enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;
KW chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW chaotropic agent; 4K-6H (904-1206)st-4K; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .15
FT /note= "4-Lys/His tag"
FT Protein 16. .302
FT /note= "Truncated NS2/3 protease"
FT Peptide 319. .334
FT /note= "Streptavidin/4-Lys tag"
XX
PN WO200248375-A2.
XX
PD 20-JUN-2002.
XX
PF 13-DEC-2001; 2001WO-CA001796.
XX
PR 15-DEC-2000; 2000US-0256031P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
XX
DR WPI; 2002-599511/64.
DR N-PSDB; ABK90407.
XX
PT Novel polypeptide for screening inhibitors of non-structural proteases
PT useful as therapeutic agents against hepatitis C virus, comprises full
PT length non-structural protease, or its truncation.
XX
PS Claim 39; Fig 9B; 67pp; English.
XX
CC The invention relates to an isolated polypeptide consisting of a full-

CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC residue amino acid 810 to 906, or having a minimal amino acid sequence
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC NS2/3 protease. Also included are (1) a composition (C) comprising an
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC its truncation or a mutated sequence, where the protease is in a solution
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC protease, involving isolating the protease in the presence of a
CC chaotropic agent, refolding the isolated protease by contacting it with a
CC reducing agent, and LDAO in the presence of reduced concentration of the
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC protease, involving diluting refolded inactive NS2/3 protease in a medium
CC containing an activation detergent to induce auto-cleavage of the NS2/3
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC protease, involving incubating the active NS2/3 protease produced by M2
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC cleavage products or their fragments, and measuring the presence or
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC active NS2/3 protease, involving carrying out M3 in the presence of, or
CC absence of the potential inhibitor, comparing the amount of uncleaved
CC NS2/3 protease, cleavage products or their fragments. The protease is
CC useful for detailed biochemical characterisation of the enzymes and in
CC the development of in vitro assays for screening novel inhibitors of
CC NS2/3 protease which are useful as therapeutic agents against HCV
CC infection (which causes chronic liver disease, cirrhosis and end-stage
CC liver disease. M1 is useful for high level production of protease. The
CC present sequence represents the NS2/3 truncation mutant 4K-6H (904-
CC 1206)st-4K comprising a truncated NS2/3 protein with a four Lys/six His N
CC -terminal tag, a C-terminal streptavidin tag and C-terminal four Lys tag
XX
SQ Sequence 334 AA;

Query Match 100.0%; Score 1771; DB 5; Length 334;
Best Local Similarity 100.0%; Pred. No. 5.4e-164;
Matches 334; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MKKKKLEHHHHHHTSAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKKKKLEHHHHHHTSAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTY 60

Qy         61 VYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGR 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGR 120

Qy        121 EILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQS 180

Qy        181 FLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSS 240

Qy        241 DLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 DLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRG 300

Qy        301 VAKAVDFIPVESMETTMRTSSAWRHPQFGGKKKK 334
              ||||||||||||||||||||||||||||||||||
Db        301 VAKAVDFIPVESMETTMRTSSAWRHPQFGGKKKK 334
```

RESULT 2
ABG32181
ID ABG32181 standard; protein; 409 AA.
XX
AC ABG32181;
XX
DT 05-NOV-2002 (first entry)
XX
DE HCV protease NS2/3 (810-1206).
XX
KW HCV; enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;
KW chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW chaotropic agent; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 398. .409
FT /note= "Streptavidin tag"
XX
PN WO200248375-A2.
XX
PD 20-JUN-2002.
XX
PF 13-DEC-2001; 2001WO-CA001796.
XX
PR 15-DEC-2000; 2000US-0256031P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
XX
DR WPI; 2002-599511/64.
DR N-PSDB; ABK90406.
XX
PT Novel polypeptide for screening inhibitors of non-structural proteases
PT useful as therapeutic agents against hepatitis C virus, comprises full
PT length non-structural protease, or its truncation.
XX
PS Claim 42; Fig 1B; 67pp; English.
XX
CC The invention relates to an isolated polypeptide consisting of a full-
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC residue amino acid 810 to 906, or having a minimal amino acid sequence
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC NS2/3 protease. Also included are (1) a composition (C) comprising an
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC its truncation or a mutated sequence, where the protease is in a solution
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC protease, involving isolating the protease in the presence of a
CC chaotropic agent, refolding the isolated protease by contacting it with a
CC reducing agent, and LDAO in the presence of reduced concentration of the
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC protease, involving diluting refolded inactive NS2/3 protease in a medium
CC containing an activation detergent to induce auto-cleavage of the NS2/3
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC protease, involving incubating the active NS2/3 protease produced by M2
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC cleavage products or their fragments, and measuring the presence or
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC active NS2/3 protease, involving carrying out M3 in the presence of, or
CC absence of the potential inhibitor, comparing the amount of uncleaved
CC NS2/3 protease, cleavage products or their fragments. The protease is
CC useful for detailed biochemical characterisation of the enzymes and in
CC the development of in vitro assays for screening novel inhibitors of
CC NS2/3 protease which are useful as therapeutic agents against HCV
CC infection (which causes chronic liver disease, cirrhosis and end-stage
CC liver disease. M1 is useful for high level production of protease. The
CC present sequence represents the NS2/3 (810-1206) protein, which has a C-
CC terminal streptavividin tag
XX
SQ Sequence 409 AA;

Query Match 93.7%; Score 1660; DB 5; Length 409;
Best Local Similarity 100.0%; Pred. No. 4.9e-153;

```
  Matches  315;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         95 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 154

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        155 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 214

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        215 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 274

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        275 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 334

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        335 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 394

Qy        316 TMRTSSAWRHPQFGG 330
              |||||||||||||||
Db        395 TMRTSSAWRHPQFGG 409


RESULT 3
ABG32183
ID   ABG32183 standard; protein; 303 AA.
XX
AC   ABG32183;
XX
DT   05-NOV-2002  (first entry)
XX
DE   HCV protease NS2/3 truncation mutant 904-1206.
XX
KW   HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;
KW   chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW   hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW   chaotropic agent; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
PN   WO200248375-A2.
XX
PD   20-JUN-2002.
XX
PF   13-DEC-2001; 2001WO-CA001796.
XX
PR   15-DEC-2000; 2000US-0256031P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Thibeault D,  Lamarre D,  Maurice R,  Pilote L,  Pause A;
XX
DR   WPI; 2002-599511/64.
XX
PT   Novel polypeptide for screening inhibitors of non-structural proteases
PT   useful as therapeutic agents against hepatitis C virus, comprises full
PT   length non-structural protease, or its truncation.
XX
PS   Claim 39; Page 58-59; 67pp; English.
XX
CC   The invention relates to an isolated polypeptide consisting of a full-
CC   length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC   to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC   residue amino acid 810 to 906, or having a minimal amino acid sequence
CC   from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC   NS2/3 protease. Also included are (1) a composition (C) comprising an
CC   isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC   its truncation or a mutated sequence, where the protease is in a solution
CC   comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC   to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC   appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC   protease, involving isolating the protease in the presence of a
CC   chaotropic agent, refolding the isolated protease by contacting it with a
CC   reducing agent, and LDAO in the presence of reduced concentration of the
CC   chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC   protease, involving diluting refolded inactive NS2/3 protease in a medium
CC   containing an activation detergent to induce auto-cleavage of the NS2/3
CC   protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC   protease, involving incubating the active NS2/3 protease produced by M2
CC   for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC   cleavage products or their fragments, and measuring the presence or
CC   absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC   ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC   active NS2/3 protease, involving carrying out M3 in the presence of, or
CC   absence of the potential inhibitor, comparing the amount of uncleaved
CC   NS2/3 protease, cleavage products or their fragments. The protease is
CC   useful for detailed biochemical characterisation of the enzymes and in
CC   the development of in vitro assays for screening novel inhibitors of
CC   NS2/3 protease which are useful as therapeutic agents against HCV
CC   infection (which causes chronic liver disease, cirrhosis and end-stage
CC   liver disease. M1 is useful for high level production of protease. The
CC   present sequence represents the NS2/3 truncation mutant 904-1206
CC   (numbered relative to the full length NS2/3 protein)
XX
SQ   Sequence 303 AA;

  Query Match           89.7%;  Score 1589;  DB 5;  Length 303;
  Best Local Similarity  100.0%;  Pred. No. 2.7e-146;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300

Qy        316 TMR 318
              |||
Db        301 TMR 303


RESULT 4
ABG32187
ID   ABG32187 standard; protein; 341 AA.
XX
AC   ABG32187;
XX
DT   05-NOV-2002  (first entry)
XX
DE   HCV protease NS2/3 truncation mutant 866-1206.
XX
KW   HCV; enzyme; protease; NS2/3 (866-1206); hepatitis C virus infection;
KW   chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW   hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW   chaotropic agent; mutant; mutein.
XX
OS   Hepatitis C virus.
```

OS Synthetic.
XX
PN WO200248375-A2.
XX
PD 20-JUN-2002.
XX
PF 13-DEC-2001; 2001WO-CA001796.
XX
PR 15-DEC-2000; 2000US-0256031P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
XX
DR WPI; 2002-599511/64.
XX
PT Novel polypeptide for screening inhibitors of non-structural proteases
PT useful as therapeutic agents against hepatitis C virus, comprises full
PT length non-structural protease, or its truncation.
XX
PS Claim 41; Page 62-63; 67pp; English.
XX
CC The invention relates to an isolated polypeptide consisting of a full-
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC residue amino acid 810 to 906, or having a minimal amino acid sequence
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC NS2/3 protease. Also included are (1) a composition (C) comprising an
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC its truncation or a mutated sequence, where the protease is in a solution
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC protease, involving isolating the protease in the presence of a
CC chaotropic agent, refolding the isolated protease by contacting it with a
CC reducing agent, and LDAO in the presence of reduced concentration of the
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC protease, involving diluting refolded inactive NS2/3 protease in a medium
CC containing an activation detergent to induce auto-cleavage of the NS2/3
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC protease, involving incubating the active NS2/3 protease produced by M2
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC cleavage products or their fragments, and measuring the presence or
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC active NS2/3 protease, involving carrying out M3 in the presence of, or
CC absence of the potential inhibitor, comparing the amount of uncleaved
CC NS2/3 protease, cleavage products or their fragments. The protease is
CC useful for detailed biochemical characterisation of the enzymes and in
CC the development of in vitro assays for screening novel inhibitors of
CC NS2/3 protease which are useful as therapeutic agents against HCV
CC infection (which causes chronic liver disease, cirrhosis and end-stage
CC liver disease. M1 is useful for high level production of protease. The
CC present sequence represents the NS2/3 truncation mutant 866-1206
CC (numbered relative to the full length NS2/3 protein)
XX
SQ Sequence 341 AA;

Query Match 89.7%; Score 1589; DB 5; Length 341;
Best Local Similarity 100.0%; Pred. No. 3.2e-146;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         39 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 98

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 158

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 218

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 278

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 338

Qy        316 TMR 318
              |||
Db        339 TMR 341
```

RESULT 5
ABG32186
ID ABG32186 standard; protein; 352 AA.
XX
AC ABG32186;
XX
DT 05-NOV-2002 (first entry)
XX
DE HCV protease NS2/3 truncation mutant 855-1206.
XX
KW HCV; enzyme; protease; NS2/3 (855-1206); hepatitis C virus infection;
KW chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW chaotropic agent; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO200248375-A2.
XX
PD 20-JUN-2002.
XX
PF 13-DEC-2001; 2001WO-CA001796.
XX
PR 15-DEC-2000; 2000US-0256031P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
XX
DR WPI; 2002-599511/64.
XX
PT Novel polypeptide for screening inhibitors of non-structural proteases
PT useful as therapeutic agents against hepatitis C virus, comprises full
PT length non-structural protease, or its truncation.
XX
PS Claim 41; Page 61-62; 67pp; English.
XX
CC The invention relates to an isolated polypeptide consisting of a full-
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC residue amino acid 810 to 906, or having a minimal amino acid sequence
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC NS2/3 protease. Also included are (1) a composition (C) comprising an
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC its truncation or a mutated sequence, where the protease is in a solution
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC protease, involving isolating the protease in the presence of a
CC chaotropic agent, refolding the isolated protease by contacting it with a
CC reducing agent, and LDAO in the presence of reduced concentration of the
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC protease, involving diluting refolded inactive NS2/3 protease in a medium
CC containing an activation detergent to induce auto-cleavage of the NS2/3
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC protease, involving incubating the active NS2/3 protease produced by M2
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC cleavage products or their fragments, and measuring the presence or

```
CC   absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC   ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC   active NS2/3 protease, involving carrying out M3 in the presence of, or
CC   absence of the potential inhibitor, comparing the amount of uncleaved
CC   NS2/3 protease, cleavage products or their fragments. The protease is
CC   useful for detailed biochemical characterisation of the enzymes and in
CC   the development of in vitro assays for screening novel inhibitors of
CC   NS2/3 protease which are useful as therapeutic agents against HCV
CC   infection (which causes chronic liver disease, cirrhosis and end-stage
CC   liver disease. M1 is useful for high level production of protease. The
CC   present sequence represents the NS2/3 truncation mutant 855-1206
CC   (numbered relative to the full length NS2/3 protein)
XX
SQ   Sequence 352 AA;

  Query Match             89.7%;  Score 1589;  DB 5;  Length 352;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-146;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          50 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 109

Qy          76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         110 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 169

Qy         136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         170 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 229

Qy         196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         230 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 289

Qy         256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         290 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 349

Qy         316 TMR 318
               |||
Db         350 TMR 352


RESULT 6
ABG32185
ID   ABG32185 standard; protein; 380 AA.
XX
AC   ABG32185;
XX
DT   05-NOV-2002  (first entry)
XX
DE   HCV protease NS2/3 truncation mutant 827-1206.
XX
KW   HCV; enzyme; protease; NS2/3 (827-1206); hepatitis C virus infection;
KW   chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW   hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW   chaotropic agent; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
PN   WO200248375-A2.
XX
PD   20-JUN-2002.
XX
PF   13-DEC-2001; 2001WO-CA001796.
XX
PR   15-DEC-2000; 2000US-0256031P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Thibeault D,  Lamarre D,  Maurice R,  Pilote L,  Pause A;
XX
DR   WPI; 2002-599511/64.
XX
PT   Novel polypeptide for screening inhibitors of non-structural proteases
PT   useful as therapeutic agents against hepatitis C virus, comprises full
PT   length non-structural protease, or its truncation.
XX
PS   Claim 41; Page 60-61; 67pp; English.
XX
CC   The invention relates to an isolated polypeptide consisting of a full-
CC   length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC   to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC   residue amino acid 810 to 906, or having a minimal amino acid sequence
CC   from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC   NS2/3 protease. Also included are (1) a composition (C) comprising an
CC   isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC   its truncation or a mutated sequence, where the protease is in a solution
CC   comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC   to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC   appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC   protease, involving isolating the protease in the presence of a
CC   chaotropic agent, refolding the isolated protease by contacting it with a
CC   reducing agent, and LDAO in the presence of reduced concentration of the
CC   chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC   protease, involving diluting refolded inactive NS2/3 protease in a medium
CC   containing an activation detergent to induce auto-cleavage of the NS2/3
CC   protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC   protease, involving incubating the active NS2/3 protease produced by M2
CC   for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC   cleavage products or their fragments, and measuring the presence or
CC   absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC   ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC   active NS2/3 protease, involving carrying out M3 in the presence of, or
CC   absence of the potential inhibitor, comparing the amount of uncleaved
CC   NS2/3 protease, cleavage products or their fragments. The protease is
CC   useful for detailed biochemical characterisation of the enzymes and in
CC   the development of in vitro assays for screening novel inhibitors of
CC   NS2/3 protease which are useful as therapeutic agents against HCV
CC   infection (which causes chronic liver disease, cirrhosis and end-stage
CC   liver disease. M1 is useful for high level production of protease. The
CC   present sequence represents the NS2/3 truncation mutant 827-1206
CC   (numbered relative to the full length NS2/3 protein)
XX
SQ   Sequence 380 AA;

  Query Match             89.7%;  Score 1589;  DB 5;  Length 380;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-146;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          78 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 137

Qy          76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         138 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 197

Qy         136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         198 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 257

Qy         196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         258 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 317

Qy         256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         318 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 377

Qy         316 TMR 318
               |||
Db         378 TMR 380
```

```
RESULT 7
ABG32184
ID   ABG32184 standard; protein; 393 AA.
XX
AC   ABG32184;
XX
DT   05-NOV-2002  (first entry)
XX
DE   HCV protease NS2/3 truncation mutant 815-1206.
XX
KW   HCV; enzyme; protease; NS2/3 (815-1206); hepatitis C virus infection;
KW   chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW   hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW   chaotropic agent; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
PN   WO200248375-A2.
XX
PD   20-JUN-2002.
XX
PF   13-DEC-2001; 2001WO-CA001796.
XX
PR   15-DEC-2000; 2000US-0256031P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Thibeault D,  Lamarre D,  Maurice R,  Pilote L,  Pause A;
XX
DR   WPI; 2002-599511/64.
XX
PT   Novel polypeptide for screening inhibitors of non-structural proteases
PT   useful as therapeutic agents against hepatitis C virus, comprises full
PT   length non-structural protease, or its truncation.
XX
PS   Claim 41; Page 59-60; 67pp; English.
XX
CC   The invention relates to an isolated polypeptide consisting of a full-
CC   length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC   to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC   residue amino acid 810 to 906, or having a minimal amino acid sequence
CC   from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC   NS2/3 protease. Also included are (1) a composition (C) comprising an
CC   isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC   its truncation or a mutated sequence, where the protease is in a solution
CC   comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC   to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC   appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC   protease, involving isolating the protease in the presence of a
CC   chaotropic agent, refolding the isolated protease by contacting it with a
CC   reducing agent, and LDAO in the presence of reduced concentration of the
CC   chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC   protease, involving diluting refolded inactive NS2/3 protease in a medium
CC   containing an activation detergent to induce auto-cleavage of the NS2/3
CC   protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC   protease, involving incubating the active NS2/3 protease produced by M2
CC   for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC   cleavage products or their fragments, and measuring the presence or
CC   absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC   ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC   active NS2/3 protease, involving carrying out M3 in the presence of, or
CC   absence of the potential inhibitor, comparing the amount of uncleaved
CC   NS2/3 protease, cleavage products or their fragments. The protease is
CC   useful for detailed biochemical characterisation of the enzymes and in
CC   the development of in vitro assays for screening novel inhibitors of
CC   NS2/3 protease which are useful as therapeutic agents against HCV
CC   infection (which causes chronic liver disease, cirrhosis and end-stage
CC   liver disease. M1 is useful for high level production of protease. The
CC   present sequence represents the NS2/3 truncation mutant 815-1206
CC   (numbered relative to the full length NS2/3 protein)
XX
SQ   Sequence 393 AA;

  Query Match             89.7%;  Score 1589;  DB 5;  Length 393;
  Best Local Similarity   100.0%;  Pred. No. 4e-146;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         91 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 150

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        151 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 210

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 270

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        271 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 330

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        331 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 390

Qy        316 TMR 318
              |||
Db        391 TMR 393


RESULT 8
ABG32191
ID   ABG32191 standard; protein; 303 AA.
XX
AC   ABG32191;
XX
DT   05-NOV-2002  (first entry)
XX
DE   HCV protease NS2/3 truncation 904-1206/Cys993Ala.
XX
KW   HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;
KW   chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW   hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW   chaotropic agent; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 90
FT                   /note= "Wild-type Cys substituted by Ala"
XX
PN   WO200248375-A2.
XX
PD   20-JUN-2002.
XX
PF   13-DEC-2001; 2001WO-CA001796.
XX
PR   15-DEC-2000; 2000US-0256031P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Thibeault D,  Lamarre D,  Maurice R,  Pilote L,  Pause A;
XX
DR   WPI; 2002-599511/64.
XX
PT   Novel polypeptide for screening inhibitors of non-structural proteases
PT   useful as therapeutic agents against hepatitis C virus, comprises full
PT   length non-structural protease, or its truncation.
XX
PS   Disclosure; Page 65-66; 67pp; English.
XX
```

```
CC   The invention relates to an isolated polypeptide consisting of a full-
CC   length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC   to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC   residue amino acid 810 to 906, or having a minimal amino acid sequence
CC   from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC   NS2/3 protease. Also included are (1) a composition (C) comprising an
CC   isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC   its truncation or a mutated sequence, where the protease is in a solution
CC   comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC   to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC   appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC   protease, involving isolating the protease in the presence of a
CC   chaotropic agent, refolding the isolated protease by contacting it with a
CC   reducing agent, and LDAO in the presence of reduced concentration of the
CC   chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC   protease, involving diluting refolded inactive NS2/3 protease in a medium
CC   containing an activation detergent to induce auto-cleavage of the NS2/3
CC   protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC   protease, involving incubating the active NS2/3 protease produced by M2
CC   for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC   cleavage products or their fragments, and measuring the presence or
CC   absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC   ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC   active NS2/3 protease, involving carrying out M3 in the presence of, or
CC   absence of the potential inhibitor, comparing the amount of uncleaved
CC   NS2/3 protease, cleavage products or their fragments. The protease is
CC   useful for detailed biochemical characterisation of the enzymes and in
CC   the development of in vitro assays for screening novel inhibitors of
CC   NS2/3 protease which are useful as therapeutic agents against HCV
CC   infection (which causes chronic liver disease, cirrhosis and end-stage
CC   liver disease. M1 is useful for high level production of protease. The
CC   present sequence represents the NS2/3 truncation 904-1206 mutant
CC   Cys993Ala (numbered relative to the full length NS2/3 protein) a mutant
CC   devoid of autocatalytic activity
XX
SQ   Sequence 303 AA;

  Query Match             89.2%;  Score 1580;  DB 5;  Length 303;
  Best Local Similarity   99.7%;  Pred. No. 2.1e-145;
  Matches  302;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db         61 LRDLAVAVEPVIFSDMEVKIITWGADTAAAGDIISGLPVSARRGREILLGPADNFEGQGW 120

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300

Qy        316 TMR 318
              |||
Db        301 TMR 303


RESULT 9
ABG32189
ID   ABG32189 standard; protein; 303 AA.
XX
AC   ABG32189;
XX
DT   05-NOV-2002  (first entry)
XX
DE   HCV protease NS2/3 truncation 904-1206/His952Ala.
XX
KW   HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;
KW   chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW   hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW   chaotropic agent; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 49
FT                   /note= "Wild-type His substituted by Ala"
XX
PN   WO200248375-A2.
XX
PD   20-JUN-2002.
XX
PF   13-DEC-2001; 2001WO-CA001796.
XX
PR   15-DEC-2000; 2000US-0256031P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Thibeault D,  Lamarre D,  Maurice R,  Pilote L,  Pause A;
XX
DR   WPI; 2002-599511/64.
XX
PT   Novel polypeptide for screening inhibitors of non-structural proteases
PT   useful as therapeutic agents against hepatitis C virus, comprises full
PT   length non-structural protease, or its truncation.
XX
PS   Example 7; Fig 8; 67pp; English.
XX
CC   The invention relates to an isolated polypeptide consisting of a full-
CC   length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC   to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC   residue amino acid 810 to 906, or having a minimal amino acid sequence
CC   from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC   NS2/3 protease. Also included are (1) a composition (C) comprising an
CC   isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC   its truncation or a mutated sequence, where the protease is in a solution
CC   comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC   to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC   appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC   protease, involving isolating the protease in the presence of a
CC   chaotropic agent, refolding the isolated protease by contacting it with a
CC   reducing agent, and LDAO in the presence of reduced concentration of the
CC   chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC   protease, involving diluting refolded inactive NS2/3 protease in a medium
CC   containing an activation detergent to induce auto-cleavage of the NS2/3
CC   protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC   protease, involving incubating the active NS2/3 protease produced by M2
CC   for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC   cleavage products or their fragments, and measuring the presence or
CC   absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC   ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC   active NS2/3 protease, involving carrying out M3 in the presence of, or
CC   absence of the potential inhibitor, comparing the amount of uncleaved
CC   NS2/3 protease, cleavage products or their fragments. The protease is
CC   useful for detailed biochemical characterisation of the enzymes and in
CC   the development of in vitro assays for screening novel inhibitors of
CC   NS2/3 protease which are useful as therapeutic agents against HCV
CC   infection (which causes chronic liver disease, cirrhosis and end-stage
CC   liver disease. M1 is useful for high level production of protease. The
CC   present sequence represents the NS2/3 truncation 904-1206 mutant
CC   His952Ala (numbered relative to the full length NS2/3 protein) a mutant
CC   devoid of autocatalytic activity
XX
SQ   Sequence 303 AA;

  Query Match             89.2%;  Score 1579;  DB 5;  Length 303;
```

```
  Best Local Similarity   99.7%;  Pred. No. 2.6e-145;
  Matches  302;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDALTPLQDWAHAG 60

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300

Qy        316 TMR 318
              |||
Db        301 TMR 303


RESULT 10
ABG32190
ID   ABG32190 standard; protein; 301 AA.
XX
AC   ABG32190;
XX
DT   05-NOV-2002  (first entry)
XX
DE   HCV protease NS2/3 truncation 904-1206/deltaLeu1026-Ala1027.
XX
KW   HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;
KW   chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW   hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW   chaotropic agent; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 122. .123
FT                   /note= "Wild-type Leu-Leu-Ala-Pro substituted by Leu-Pro"
XX
PN   WO200248375-A2.
XX
PD   20-JUN-2002.
XX
PF   13-DEC-2001; 2001WO-CA001796.
XX
PR   15-DEC-2000; 2000US-0256031P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Thibeault D,  Lamarre D,  Maurice R,  Pilote L,  Pause A;
XX
DR   WPI; 2002-599511/64.
XX
PT   Novel polypeptide for screening inhibitors of non-structural proteases
PT   useful as therapeutic agents against hepatitis C virus, comprises full
PT   length non-structural protease, or its truncation.
XX
PS   Example 7; Page 64-65; 67pp; English.
XX
CC   The invention relates to an isolated polypeptide consisting of a full-
CC   length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC   to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC   residue amino acid 810 to 906, or having a minimal amino acid sequence
CC   from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC   NS2/3 protease. Also included are (1) a composition (C) comprising an
CC   isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC   its truncation or a mutated sequence, where the protease is in a solution
CC   comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC   to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC   appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC   protease, involving isolating the protease in the presence of a
CC   chaotropic agent, refolding the isolated protease by contacting it with a
CC   reducing agent, and LDAO in the presence of reduced concentration of the
CC   chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC   protease, involving diluting refolded inactive NS2/3 protease in a medium
CC   containing an activation detergent to induce auto-cleavage of the NS2/3
CC   protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC   protease, involving incubating the active NS2/3 protease produced by M2
CC   for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC   cleavage products or their fragments, and measuring the presence or
CC   absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC   ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC   active NS2/3 protease, involving carrying out M3 in the presence of, or
CC   absence of the potential inhibitor, comparing the amount of uncleaved
CC   NS2/3 protease, cleavage products or their fragments. The protease is
CC   useful for detailed biochemical characterisation of the enzymes and in
CC   the development of in vitro assays for screening novel inhibitors of
CC   NS2/3 protease which are useful as therapeutic agents against HCV
CC   infection (which causes chronic liver disease, cirrhosis and end-stage
CC   liver disease. M1 is useful for high level production of protease. The
CC   present sequence represents the NS2/3 truncation 904-1206 mutant
CC   deltaLeu1026-Ala1027 (numbered relative to the full length NS2/3 protein)
CC   a mutant devoid of autocatalytic activity
XX
SQ   Sequence 301 AA;

  Query Match             88.7%;  Score 1570;  DB 5;  Length 301;
  Best Local Similarity   99.3%;  Pred. No. 1.9e-144;
  Matches  301;  Conservative    0;  Mismatches    0;  Indels    2;  Gaps    1;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RL--PITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 178

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 238

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 298

Qy        316 TMR 318
              |||
Db        299 TMR 301


RESULT 11
ABG32188
ID   ABG32188 standard; protein; 292 AA.
XX
AC   ABG32188;
XX
DT   05-NOV-2002  (first entry)
XX
DE   HCV protease NS2/3 truncation mutant 915-1206.
XX
```

```
KW   HCV; enzyme; protease; NS2/3 (915-1206); hepatitis C virus infection;
KW   chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW   hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW   chaotropic agent; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
PN   WO200248375-A2.
XX
PD   20-JUN-2002.
XX
PF   13-DEC-2001; 2001WO-CA001796.
XX
PR   15-DEC-2000; 2000US-0256031P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Thibeault D,  Lamarre D,  Maurice R,  Pilote L,  Pause A;
XX
DR   WPI; 2002-599511/64.
XX
PT   Novel polypeptide for screening inhibitors of non-structural proteases
PT   useful as therapeutic agents against hepatitis C virus, comprises full
PT   length non-structural protease, or its truncation.
XX
PS   Claim 41; Page 63; 67pp; English.
XX
CC   The invention relates to an isolated polypeptide consisting of a full-
CC   length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC   to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC   residue amino acid 810 to 906, or having a minimal amino acid sequence
CC   from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC   NS2/3 protease. Also included are (1) a composition (C) comprising an
CC   isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC   its truncation or a mutated sequence, where the protease is in a solution
CC   comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC   to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC   appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC   protease, involving isolating the protease in the presence of a
CC   chaotropic agent, refolding the isolated protease by contacting it with a
CC   reducing agent, and LDAO in the presence of reduced concentration of the
CC   chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC   protease, involving diluting refolded inactive NS2/3 protease in a medium
CC   containing an activation detergent to induce auto-cleavage of the NS2/3
CC   protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC   protease, involving incubating the active NS2/3 protease produced by M2
CC   for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC   cleavage products or their fragments, and measuring the presence or
CC   absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC   ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC   active NS2/3 protease, involving carrying out M3 in the presence of, or
CC   absence of the potential inhibitor, comparing the amount of uncleaved
CC   NS2/3 protease, cleavage products or their fragments. The protease is
CC   useful for detailed biochemical characterisation of the enzymes and in
CC   the development of in vitro assays for screening novel inhibitors of
CC   NS2/3 protease which are useful as therapeutic agents against HCV
CC   infection (which causes chronic liver disease, cirrhosis and end-stage
CC   liver disease. M1 is useful for high level production of protease. The
CC   present sequence represents the NS2/3 truncation mutant 915-1206
CC   (numbered relative to the full length NS2/3 protein)
XX
SQ   Sequence 292 AA;

  Query Match             86.5%;  Score 1532;  DB 5;  Length 292;
  Best Local Similarity   100.0%;  Pred. No. 9.5e-141;
  Matches  292;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 AQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPV 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPV 60

Qy         87 IFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQ 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQ 120

Qy        147 QTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPK 180

Qy        207 GPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLS 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLS 240

Qy        267 PRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 318
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 292


RESULT 12
ABG30601
ID   ABG30601 standard; protein; 2201 AA.
XX
AC   ABG30601;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #10.
XX
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 882
FT                   /label= Arg, Lys
FT   Misc-difference 2183
FT                   /note= "Wild type Met substituted by Thr"
XX
PN   WO200252015-A2.
XX
PD   04-JUL-2002.
XX
PF   20-DEC-2001; 2001WO-CA001843.
XX
PR   22-DEC-2000; 2000US-0257857P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Kukolj G,  Pause A;
XX
DR   WPI; 2002-575382/61.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
XX
PS   Claim 3; Page; 140pp; English.
XX
CC   The invention describes a self-replicating hepatitis C virus (HCV)
CC   polynucleotide molecule comprising a 5'-non translated region (NTR),
CC   where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC   region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC   replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC   potential inhibitors of HCV replication. The HCV RNA molecule is also
CC   useful for efficiently establishing cell culture replication. The self-
CC   replicating polynucleotide molecule contains a 5'-NTR, where G at
CC   position 1 is substituted for A, and therefore provides an alternative to
CC   existing systems comprising a self-replicating HCV RNA molecule that, in
CC   conjunction with mutations in the HCV non-structural region, such as the
CC   G(2042)C/R mutations, transduces and/or replicates with greater
CC   efficiency. This amino acid sequence represents a mutant of the hepatitis
CC   C virus replicon APGK12 and contains the viral protease NS2/3, protease
CC   complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
```

```
CC   This sequence does not appear in the specification but has been created
CC   from the wild type sequence shown in ABG30580 using information given in
CC   the claims of the invention
XX
SQ   Sequence 2201 AA;

  Query Match             86.4%;  Score 1531;  DB 5;  Length 2201;
  Best Local Similarity   94.7%;  Pred. No. 2.1e-139;
  Matches  288;  Conservative    9;  Mismatches    7;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              |||||||||||| |||||||||| |||||||| ||||||||||||||||||:||||||
Db         95 AGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAG 154

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||:||||| |:||||||||||||| |||||||||| |||||: |||||
Db        155 LRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQGW 214

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||:||||||||||||||||||||||||:|
Db        215 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 274

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db        275 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 334

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              |||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        335 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 394

Qy        316 TMRT 319
              |||:
Db        395 TMRS 398


RESULT 13
ABG30591
ID   ABG30591 standard; protein; 2201 AA.
XX
AC   ABG30591;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #3.
XX
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 751
FT                   /note= "Wild type Ser substituted by Gly"
FT   Misc-difference 882
FT                   /label= Arg, Lys
XX
PN   WO200252015-A2.
XX
PD   04-JUL-2002.
XX
PF   20-DEC-2001; 2001WO-CA001843.
XX
PR   22-DEC-2000; 2000US-0257857P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Kukolj G,  Pause A;
XX
DR   WPI; 2002-575382/61.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
XX
PS   Claim 3; Page; 140pp; English.
XX
CC   The invention describes a self-replicating hepatitis C virus (HCV)
CC   polynucleotide molecule comprising a 5'-non translated region (NTR),
CC   where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC   region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC   replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC   potential inhibitors of HCV replication. The HCV RNA molecule is also
CC   useful for efficiently establishing cell culture replication. The self-
CC   replicating polynucleotide molecule contains a 5'-NTR, where G at
CC   position 1 is substituted for A, and therefore provides an alternative to
CC   existing systems comprising a self-replicating HCV RNA molecule that, in
CC   conjunction with mutations in the HCV non-structural region, such as the
CC   G(2042)C/R mutations, transduces and/or replicates with greater
CC   efficiency. This amino acid sequence represents a mutant of the hepatitis
CC   C virus replicon APGK12 and contains the viral protease NS2/3, protease
CC   complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
CC   This sequence does not appear in the specification but has been created
CC   from the wild type sequence shown in ABG30580 using information given in
CC   the claims of the invention
XX
SQ   Sequence 2201 AA;

  Query Match             86.4%;  Score 1531;  DB 5;  Length 2201;
  Best Local Similarity   94.7%;  Pred. No. 2.1e-139;
  Matches  288;  Conservative    9;  Mismatches    7;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              |||||||||||| |||||||||| |||||||| ||||||||||||||||||:||||||
Db         95 AGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAG 154

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||:||||| |:||||||||||||| |||||||||| |||||: |||||
Db        155 LRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQGW 214

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||:||||||||||||||||||||||||:|
Db        215 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 274

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db        275 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 334

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              |||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        335 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 394

Qy        316 TMRT 319
              |||:
Db        395 TMRS 398


RESULT 14
ABG30600
ID   ABG30600 standard; protein; 2201 AA.
XX
AC   ABG30600;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #9.
XX
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 882
FT                   /label= Arg, Lys
```

```
FT   Misc-difference 1357
FT                   /note= "Wild type Pro substituted by Leu"
XX
PN   WO200252015-A2.
XX
PD   04-JUL-2002.
XX
PF   20-DEC-2001; 2001WO-CA001843.
XX
PR   22-DEC-2000; 2000US-0257857P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Kukolj G,  Pause A;
XX
DR   WPI; 2002-575382/61.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
XX
PS   Claim 3; Page; 140pp; English.
XX
CC   The invention describes a self-replicating hepatitis C virus (HCV)
CC   polynucleotide molecule comprising a 5'-non translated region (NTR),
CC   where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC   region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC   replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC   potential inhibitors of HCV replication. The HCV RNA molecule is also
CC   useful for efficiently establishing cell culture replication. The self-
CC   replicating polynucleotide molecule contains a 5'-NTR, where G at
CC   position 1 is substituted for A, and therefore provides an alternative to
CC   existing systems comprising a self-replicating HCV RNA molecule that, in
CC   conjunction with mutations in the HCV non-structural region, such as the
CC   G(2042)C/R mutations, transduces and/or replicates with greater
CC   efficiency. This amino acid sequence represents a mutant of the hepatitis
CC   C virus replicon APGK12 and contains the viral protease NS2/3, protease
CC   complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
CC   This sequence does not appear in the specification but has been created
CC   from the wild type sequence shown in ABG30580 using information given in
CC   the claims of the invention
XX
SQ   Sequence 2201 AA;

  Query Match             86.4%;  Score 1531;  DB 5;  Length 2201;
  Best Local Similarity   94.7%;  Pred. No. 2.1e-139;
  Matches  288;  Conservative    9;  Mismatches    7;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              |||||||||||| |||||||||| |||||||| ||||||||||||||||||:||||||
Db         95 AGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAG 154

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              |||||||||||:||||| |:||||||||||||| |||||||||||| ||||||: |||||
Db        155 LRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQGW 214

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              |||||||||||||||||||||||||:|||||||||||||||||||||||||||||||:|
Db        215 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 274

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              |||||||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db        275 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 334

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              |||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        335 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 394

Qy        316 TMRT 319
              ||||
Db        395 TMRS 398
```

```
RESULT 15
ABG30581
ID   ABG30581 standard; protein; 2201 AA.
XX
AC   ABG30581;
XX
DT.  21-OCT-2002  (first entry)
XX
DE   Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #1.
XX
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; NS2/3; NS3/4; NS3; NS5B.
XX
OS   Hepatitis C virus.
XX
PN   WO200252015-A2.
XX
PD   04-JUL-2002.
XX
PF   20-DEC-2001; 2001WO-CA001843.
XX
PR   22-DEC-2000; 2000US-0257857P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Kukolj G,  Pause A;
XX
DR   WPI; 2002-575382/61.
DR   N-PSDB; ABK88573.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
XX
PS   Disclosure; Page 49-58; 140pp; English.
XX
CC   The invention describes a self-replicating hepatitis C virus (HCV)
CC   polynucleotide molecule comprising a 5'-non translated region (NTR),
CC   where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC   region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC   replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC   potential inhibitors of HCV replication. The HCV RNA molecule is also
CC   useful for efficiently establishing cell culture replication. The self-
CC   replicating polynucleotide molecule contains a 5'-NTR, where G at
CC   position 1 is substituted for A, and therefore provides an alternative to
CC   existing systems comprising a self-replicating HCV RNA molecule that, in
CC   conjunction with mutations in the HCV non-structural region, such as the
CC   G(2042)C/R mutations, transduces and/or replicates with greater
CC   efficiency. This amino acid sequence is encoded by the hepatitis C virus
CC   replicon APGK12 and contains the viral protease NS2/3, protease complex
CC   NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B
XX
SQ   Sequence 2201 AA;

  Query Match             86.4%;  Score 1531;  DB 5;  Length 2201;
  Best Local Similarity   94.7%;  Pred. No. 2.1e-139;
  Matches  288;  Conservative    9;  Mismatches    7;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              |||||||||||| |||||||||| |||||||| ||||||||||||||||||:||||||
Db         95 AGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAG 154

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              |||||||||||:||||| |:||||||||||||| |||||||||||| ||||||: |||||
Db        155 LRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQGW 214

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              |||||||||||||||||||||||||:|||||||||||||||||||||||||||||||:|
Db        215 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 274

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              |||||||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db        275 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 334
```

```
Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        335 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 394

Qy        316 TMRT 319
              |||:
Db        395 TMRS 398
```

Search completed: May 6, 2004, 09:30:43
Job time : 46.463 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:22:36 ; Search time 10.8777 Seconds
(without alignments)
2953.573 Million cell updates/sec

Title: US-10-650-585-4
Perfect score: 1771
Sequence: 1 MKKKKLEHHHHHHTSAGITK.....TTMRTSSAWRHPQFGGKKKK 334

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1533	86.6	3010	1	A45573	genome polyprotein
2	1528	86.3	3010	1	GNWVCJ	genome polyprotein
3	1515	85.5	3010	1	GNWVTW	genome polyprotein
4	1487	84.0	3010	1	S18030	genome polyprotein
5	1479	83.5	3010	1	GNWVTC	genome polyprotein
6	1404	79.3	3011	1	S40770	genome polyprotein
7	1399	79.0	3011	1	GNWVC3	genome polyprotein
8	1386	78.3	3011	1	GNWVCH	genome polyprotein
9	1235	69.7	3014	1	JC5620	genome polyprotein
10	1173	66.2	3033	1	JQ1303	genome polyprotein
11	1158	65.4	3033	1	GNWVJ8	genome polyprotein
12	398.5	22.5	3005	2	T08841	polyprotein - dour
13	343	19.4	2970	2	T08839	polyprotein - marm
14	108	6.1	353	2	G87392	conserved hypothet
15	101	5.7	600	2	B46642	DNA-directed DNA p
16	98.5	5.6	470	2	JC4098	tetracycline 6-hyd
17	97.5	5.5	1085	2	T03531	cobN protein homol
18	95.5	5.4	706	2	S33761	transferrin precur
19	94.5	5.3	716	2	G83612	hypothetical prote
20	93.5	5.3	7463	2	T36248	CDA peptide synthe
21	93	5.3	660	1	VHWWH2	structural protein
22	91	5.1	904	2	A84212	hypothetical prote
23	90.5	5.1	868	2	H81775	aconitate hydratas
24	90	5.1	659	1	B44212	structural protein
25	90	5.1	2796	2	JC4743	fatty-acid synthas
26	89.5	5.1	267	2	B83602	conserved hypothet
27	88	5.0	3414	1	GNWVNE	genome polyprotein
28	87	4.9	1057	2	I38171	hugl protein - hum
29	87	4.9	3412	1	GNWVTB	genome polyprotein
30	86.5	4.9	470	1	NMIVW8	exo-alpha-sialidas
31	86.5	4.9	990	2	S67499	glutamate synthase
32	85.5	4.8	299	2	AH3447	cytochrome-c oxida
33	85.5	4.8	348	2	H70549	probable pdhB prot
34	85	4.8	470	1	NMIV9	exo-alpha-sialidas
35	85	4.8	707	2	D84154	cadmium-transporti
36	84.5	4.8	347	2	S44167	malate dehydrogena
37	84.5	4.8	5627	2	C83339	hypothetical prote
38	84	4.7	233	2	T35594	hypothetical prote
39	84	4.7	3069	2	H70656	fatty-acid synthas
40	83.5	4.7	315	2	AG2361	hypothetical prote
41	83.5	4.7	538	2	S22409	D-alanyl-D-alanine
42	83.5	4.7	1399	2	G83112	DNA-directed RNA p
43	83	4.7	398	2	B71284	probable periplasm
44	83	4.7	2103	2	G86925	probable polyketid
45	83	4.7	4735	2	T17463	rifamycin polyketi

ALIGNMENTS

RESULT 1
A45573
genome polyprotein - hepatitis C virus (strain JT)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C;Accession: A45573
R;Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,
Virus Res. 23, 39-53, 1992
A;Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier:
A;Reference number: A45573; MUID:92295714; PMID:1318627
A;Accession: A45573
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-3010 <TAN>
A;Cross-references: GB:D11168; GB:D01171; NID:g221612; PIDN:BAA01943.1; PID:g221613
A;Experimental source: HCV-JT
A;Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBIP:106207)
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 86.6%; Score 1533; DB 1; Length 3010;
Best Local Similarity 95.1%; Pred. No. 2.7e-121;
Matches 289; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

```
Qy          16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
               | || :|||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db         904 AAITAMPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 963

Qy          76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
               |||||||||||:||||| |||||||||||||||| |||||||||||||||||: |||||
Db         964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGPADSIEGQGW 1023

Qy         136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
               |||||||||:||||||||||:||||||||||||||||||||||||||||||||||||||
Db        1024 RLLAPITAYAQQTRGLLGCIVTSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 1083

Qy         196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
               |||||||||||||||||||||||||||| |||||||:|||||||||||||||||||||
```

```
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWHAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 RRGDGRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203

Qy        316 TMRT 319
              |||:
Db       1204 TMRS 1207
```

RESULT 2
GNWVCJ
genome polyprotein - hepatitis C virus (strain J)
N;Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructura
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001
C;Accession: A39253; PS0086
R;Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimot
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
A;Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients
A;Reference number: A39253; MUID:91088550; PMID:2175903
A;Accession: A39253
A;Molecule type: genomic RNA
A;Residues: 1-3010 <KAT>
A;Cross-references: GB:D90208; NID:g221610; PIDN:BAA14233.1; PID:g221611
R;Kato, N.; Ohkoshi, S.; Shimotohno, K.
Proc. Jpn. Acad. 65B, 219-223, 1989
A;Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari
A;Reference number: PS0085
A;Accession: PS0086
A;Molecule type: genomic RNA
A;Residues: 2650-2707 <KA2>
A;Experimental source: Japanese isolate
C;Comment: The cleavage sites of this polyprotein have not been determined.
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2

Query Match 86.3%; Score 1528; DB 1; Length 3010;
Best Local Similarity 94.1%; Pred. No. 7.1e-121;
Matches 286; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

```
Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||:|||||||||||||||||| ||||||||||||||||||||||||:||||||
Db        904 AGITRVPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLRDWAHAG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||:||||| |:|||||||||||||||||||||:||||||:|  |||
Db        964 LRDLAVAVEPVVFSDMETKLITWGADTAACGDIISGLPVSARRGKEILLGPADSFGEQGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||:|||||:||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVDGEVQVLSTATQSFLATCVNGVCWTVYH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              |||||||||||||||||||||||||| |||||||||||||||||||||||:|||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWPAPPGARSMTPCTCGSSDLYLVTRHADVVPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
```

```
              |||||||||||||:|||||||||||||||| |||||||||||||||||||||||||||
Db       1144 RRGDSRGSLLSPRPISYLKGSSGGPLLCPSGHVVGIFRAAVCTRGVAKAVDFIPVESMET 1203

Qy        316 TMRT 319
              |||:
Db       1204 TMRS 1207
```

RESULT 3
GNWVTW
genome polyprotein - hepatitis C virus (strain Taiwan)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: A40244
R;Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992
A;Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
A;Reference number: A40244; MUID:92230206; PMID:1314449
A;Accession: A40244
A;Molecule type: genomic RNA
A;Residues: 1-3010 <CHE>
A;Cross-references: GB:M84754
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match 85.5%; Score 1515; DB 1; Length 3010;
Best Local Similarity 92.8%; Pred. No. 9e-120;
Matches 282; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

```
Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||:|::|||||||||||||||||| |||||||||||||||||||||||||||||| |
Db        904 AGLTRIPYFVRAQGLIRACMLVRKAAGGHYVQMALMKLAALTGTYVYDHLTPLQDWAHTG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||:||||| |||||||||||||| ||||||||||||||||: ||:||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGPADSLEGRGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              |||||||||:|||||| ||||||||||||||||||||||||||||:|||||||:|
Db       1024 RLLAPITAYAQQTRGLFGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCINGVCWTVYH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              |||||||||||||||||||||||||| || ||||:|||||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWHAPQGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 315
              |||||||||||||:|||||||||||||||| |||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPISYLKGSSGGPLLCPSGHVVGIFRAAVCTRGVAKAVDFIPVESMET 1203

Qy        316 TMRT 319
              |||:
Db       1204 TMRS 1207
```

RESULT 4
S18030
genome polyprotein - hepatitis C virus (isolate JK1)

```
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Variety: isolate JK1
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001
C;Accession: S18030; S33570; A48332; S18029
R;Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A;Description: A whole genome of hepatitis C virus cDNA was isolated from a single patie
A;Reference number: S18028
A;Accession: S18030
A;Molecule type: genomic RNA
A;Residues: 1-3010 <HON>
A;Cross-references: EMBL:X61596; NID:g59478; PIDN:CAA43793.1; PID:g59479
A;Experimental source: isolate JK1 from an individual
R;Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.
Arch. Virol. 128, 163-169, 1993
A;Title: Sequence analysis of putative structural regions of hepatitis C virus isolated
A;Reference number: A48332; MUID:93119270; PMID:8380322
A;Accession: S33570
A;Molecule type: genomic RNA
A;Residues: 1-547,'T',549-621,'V',623-624,'S',626-652,'DL',655-761,'T',763-782 <HOW>
A;Cross-references: EMBL:X61591
A;Note: this sequence is inconsistent with the nucleotide translation
A;Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320
 as Trp, and TTC for residue 771 as Ser
A;Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBIP:121748)
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (As

  Query Match             84.0%;  Score 1487;  DB 1;  Length 3010;
  Best Local Similarity   92.1%;  Pred. No. 2.1e-117;
  Matches  280;  Conservative   10;  Mismatches   14;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||:|:|||||||||:|||||||| |||||||| |||||||||||||||||:||||:|
Db        904 AGLTRVPYFVRAQGLLRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHSG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              |||| |||||:||||| |||||||||||||| |||||||||||||||||| |  |||
Db        964 LRDLVVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGPADGFREQGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              |||||||||||||||| |||:|||||||||||| |||||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLFGCIVTSLTGRDKNQVEGEAQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              |||||||||||||||||||||||||||| || |:|||| ||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPINQMYTNVDQDLVGWQAPSGAASLTPCTYGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203

Qy        316 TMRT 319
              ||||:
Db       1204 TMRS 1207


RESULT 5
GNWVTC
genome polyprotein - hepatitis C virus
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
C;Accession: A38465
R;Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.;
J. Virol. 65, 1105-1113, 1991
A;Title: Structure and organization of the hepatitis C virus genome isolated from human
A;Reference number: A38465; MUID:91140698; PMID:1847440
A;Accession: A38465
A;Molecule type: genomic RNA
A;Residues: 1-3010 <TAK>
A;Cross-references: EMBL:M58335; NID:g329770; PIDN:AAA72945.1; PID:g329771
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,22

  Query Match             83.5%;  Score 1479;  DB 1;  Length 3010;
  Best Local Similarity   91.8%;  Pred. No. 1e-116;
  Matches  279;  Conservative   12;  Mismatches   13;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||:|||||||||||| ||||||| |||||||||||||| ||||||:|:||||||:||  ||
Db        904 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWPRAG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||:||||| |||||||||||||| |||||||||:|||||: ||:|
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGRGL 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              |||||||| |||||||||||||||||  ||||||:|||||||||||||||||||||||
Db       1084 GAGSKTLAAPKGPITQMYTNVDQDLVGWPKPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              |||||||||||||||||||||||||||| ||||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPFGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        316 TMRT 319
              |||:
Db       1204 TMRS 1207


RESULT 6
S40770
genome polyprotein - hepatitis C virus
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C;Accession: S40770; PC1285
R;Okamoto, H.
submitted to the EMBL Data Library, March 1992
A;Reference number: S40770
A;Accession: S40770
A;Molecule type: genomic RNA
```

A;Residues: 1-3011 <OKA>
A;Cross-references: EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g221587
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,
Jpn. J. Exp. Med. 60, 167-177, 1990
A;Title: The 5'-terminal sequence of the hepatitis C virus genome.
A;Reference number: PC1284; MUID:91013116; PMID:2170712
A;Accession: PC1285
A;Molecule type: genomic RNA
A;Residues: 1-513 <OK2>
A;Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512
A;Experimental source: isolate HC-J1
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 79.3%; Score 1404; DB 1; Length 3011;
Best Local Similarity 84.5%; Pred. No. 2.4e-110;
Matches 257; Conservative 24; Mismatches 23; Indels 0; Gaps 0;

```
Qy        16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
             | : ||||||| |||:| | | ||  |||||||| :|| ||||||||:|||||:|||| |
Db       904 ASLLKVPYFVRVQGLLRICALARKMVGGHYVQMAIIKLGALTGTYVYNHLTPLRDWAHNG 963

Qy        76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
             |||||||||||:|| || |:|||||||||||||:|||||||:||||||||    :||
Db       964 LRDLAVAVEPVVFSQMETKLITWGADTAACGDIINGLPVSARKGREILLGPADGMVSKGW 1023

Qy       136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
             |||||||||:|||||||||||||||||||||||||||:|||| |:|||||:|||||||:|
Db      1024 RLLAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYH 1083

Qy       196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
             |||::|:| |||||: |||||||||||||| || ||||:|||||||||||||||||||||
Db      1084 GAGTRTIASPKGPVIQMYTNVDQDLVGWPAPQGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy       256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
             |||||||||||||||:|||||||||||||:|| ||||||||||||||||||||||||:||
Db      1144 RRGDSRGSLLSPRPISYLKGSSGGPLLCPAGHVVGIFRAAVCTRGVAKAVDFIPVESLET 1203

Qy       316 TMRT 319
             ||||
Db      1204 TMRS 1207
```

RESULT 7
GNWVC3
genome polyprotein - hepatitis C virus (strain HCV-1)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001
C;Accession: A39166; PQ0403; PQ0404
R;Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
A;Title: Genetic organization and diversity of the hepatitis C virus.
A;Reference number: A39166; MUID:91172826; PMID:1848704
A;Accession: A39166
A;Molecule type: mRNA
A;Residues: 1-3011 <CHO>
A;Cross-references: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874
R;Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L
J. Gen. Virol. 73, 1131-1141, 1992

A;Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
A;Reference number: PQ0393; MUID:92268871; PMID:1316939
A;Accession: PQ0403
A;Molecule type: genomic RNA
A;Residues: 1577-1633 <CHA>
A;Cross-references: DDBJ:D10128
A;Experimental source: isolates E-b16
A;Accession: PQ0404
A;Status: preliminary
A;Molecule type: genomic RNA
A;Residues: 1577-1633 <CH2>
A;Experimental source: isolates E-b17
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 79.0%; Score 1399; DB 1; Length 3011;
Best Local Similarity 84.2%; Pred. No. 6.3e-110;
Matches 256; Conservative 25; Mismatches 23; Indels 0; Gaps 0;

```
Qy        16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
             | : ||||||| |||:| | | ||  |||||||  :|| |||||||:|||||:|||| |
Db       904 ASLLKVPYFVRVQGLLRFCALARKMIGGHYVQMVIIKLGALTGTYVYNHLTPLRDWAHNG 963

Qy        76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
             |||||||||||:|| || |:|||||||||||||:|||||||||||||||||    :||
Db       964 LRDLAVAVEPVVFSQMETKLITWGADTAACGDIINGLPVSARRGREILLGPADGMVSKGW 1023

Qy       136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
             |||||||||:|||||||||||||||||||||||||||:|||| |:|||||:|||||||:|
Db      1024 RLLAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYH 1083

Qy       196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
             |||::|:| |||||: |||||||||||||| || |:|::|||||||||||||||||||||
Db      1084 GAGTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy       256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
             |||||||||||||||:|||||||||||||:|| |||||||||||||||||||||||||:||
Db      1144 RRGDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLET 1203

Qy       316 TMRT 319
             ||||
Db      1204 TMRS 1207
```

RESULT 8
GNWVCH
genome polyprotein - hepatitis C virus (strain H)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: A36814; A41546
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A;Description: Genomic structure of the human prototype strain H of hepatitis C virus: co
A;Reference number: A36814
A;Accession: A36814
A;Molecule type: genomic RNA
A;Residues: 1-3011 <INC>

```
A;Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A;Title: Genomic structure of the human prototype strain H of hepatitis C virus: compari
A;Reference number: A41546; MUID:92052256; PMID:1658800
A;Contents: annotation
A;Note: neither amino acid nor nucleotide sequence is given
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

  Query Match             78.3%;  Score 1386;  DB 1;  Length 3011;
  Best Local Similarity   83.6%;  Pred. No. 8e-109;
  Matches  254;  Conservative   27;  Mismatches   23;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              | : ||||||| |||:| | | || ||||||||| :|| ||||| ||:|| ||:|||| |
Db        904 ASLLKVPYFVRVQGLLRICALARKIAGGHYVQMAIIKLGALTGTCVYNHLAPLRDWAHNG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              |||||||||||:|| || |:|||||||||||||:|||||||:||||||||    :||
Db        964 LRDLAVAVEPVVFSRMETKLITWGADTAACGDIINGLPVSARRGQEILLGPADGMVSKGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              |||||||||:|||||||||||||||||||||||||||||:||||||:|||||:|||||||:|
Db       1024 RLLAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTATQTFLATCINGVCWTVYH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              |||::|:| ||||: | |||||||||| || |:||:|||||||||||||||||||||||
Db       1084 GAGTRTIASPKGPVIQTYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||:|||||||||||||:||||||:|||||||||||||||||||:||
Db       1144 RRGDSRGSLLSPRPISYLKGSSGGPLLCPTGHAVGLFRAAVCTRGVAKAVDFIPVENLET 1203

Qy        316 TMRT 319
              |||:
Db       1204 TMRS 1207


RESULT 9
JC5620
genome polyprotein - hepatitis C virus (isolate EUH1480)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C;Accession: JC5620
R;Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
Biochem. Biophys. Res. Commun. 236, 44-49, 1997
A;Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant
A;Reference number: JC5620; MUID:97366593; PMID:9223423
A;Accession: JC5620
A;Molecule type: mRNA
A;Residues: 1-3014 <CHA>
A;Cross-references: GB:Y13184
A;Experimental source: genotype 5a, which predominates in South Africa
A;Note: the translation of the nucleotide sequence is not complete in this paper
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;384-408/Region: hypervariable #status predicted
F;390-730/Product: nonstructural protein NS1 #status predicted <NS1>
F;731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
F;1008-1616/Product: hepacivirin #status predicted <NS3>
F;1231-1238/Region: nucleotide-binding motif A (P-loop)
F;1313-1318/Region: nucleotide-binding motif B
F;1317-1320/Region: DEXH motif
F;1617-1863/Product: nonstructural protein NS4a #status predicted <N4A>
F;1864-2014/Product: nonstructural protein NS4b #status predicted <N4B>
F;2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
F;2210-2249/Region: interferon sensitivity determining #status predicted

  Query Match             69.7%;  Score 1235;  DB 1;  Length 3014;
  Best Local Similarity   72.5%;  Pred. No. 5.3e-96;
  Matches  219;  Conservative   44;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         18 ITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAGLR 77
              :||||||:||: |:| |:| :    | ||| | : |  |||||:|||| |::||| :|||
Db        907 LTKVPYFLRARALLRLCLLAKHLVYGKYVQAALLHLGRLTGTYIYDHLAPMKDWAASGLR 966

Qy         78 DLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGWRL 137
              :| || ||::|| || |:|||||||||||:|:: ||||||||||| |||||: :  ||||
Db        967 ELTVATEPIVFSAMETKVITWGADTAACGNILAVLPVSARRGREIFLGPADDIKTSGWRL 1026

Qy        138 LAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFHGA 197
              |||||||:|||||:|| |: |||||||| |||||  :|||||:||  |:||| ||:||||
Db       1027 LAPITAYAQQTRGVLGAIVLSLTGRDKNEAEGEVQFLSTATQTFLGICINGVMWTLFHGA 1086

Qy        198 GSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVRRR 257
              |||||||||||: |||||||:||||| :|||  |:| |||||:|||||||||||||| |||
Db       1087 GSKTLAGPKGPVVQMYTNVDKDLVGWPSPPGKGSLTRCTCGSADLYLVTRHADVIPARRR 1146

Qy        258 GDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTM 317
              ||:| |||||||:|||||||||::||||| ||:|||||||||||::|:|||::||||
Db       1147 GDTRASLLSPRPISYLKGSSGGPIMCPSGHVVGVFRAAVCTRGVAKALEFVPVENLETTM 1206

Qy        318 RT 319
              |:
Db       1207 RS 1208


RESULT 10
JQ1303
genome polyprotein - hepatitis C virus (isolate HC-J6)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
C;Accession: JQ1303
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.;
J. Gen. Virol. 72, 2697-2704, 1991
A;Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a huma
A;Reference number: JQ1303; MUID:92044440; PMID:1658196
A;Accession: JQ1303
A;Molecule type: genomic RNA
A;Residues: 1-3033 <OKA>
A;Cross-references: GB:D00944; NID:g221650; PIDN:BAA00792.1; PID:g221651
A;Experimental source: isolate HC-J6 from a Japanese individual
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transme
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F;1011-1619/Product: hepacivirin #status predicted <NS3>
F;1316-1321/Region: nucleotide-binding motif B
F;1320-1323/Region: DEXH motif
F;1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>
F;1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>
F;2018-3033/Product: nonstructural protein NS5 #status predicted <NO5>
```

F;196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28

Query Match 66.2%; Score 1173; DB 1; Length 3033;
Best Local Similarity 69.2%; Pred. No. 9.8e-91;
Matches 209; Conservative 45; Mismatches 48; Indels 0; Gaps 0;

```
Qy        18 ITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAGLR 77
             :|:|||||||  |:| | :||  ||| ||||  : |   ||||:||||||: |||  |||
Db       910 LTRVPYFVRAHALLRMCTMVRHLAGGRYVQMVLLALGRWTGTYIYDHLTPMSDWAANGLR 969

Qy        78 DLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGWRL 137
             ||||||||:|||  || |:| |||:|||||||: ||||||| |||:|||||| :  :|| |
Db       970 DLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPADGYTSKGWSL 1029

Qy       138 LAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFHGA 197
             |||||||:|||||||| |: |:||||| :  ||:||:|| ||||| | ::|| |||:|||
Db      1030 LAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTTISGVLWTVYHGA 1089

Qy       198 GSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVRRR 257
             |:||||| :||:||||:: : ||||| :||| :|: |||||: |||||||:|||| |||
Db      1090 GNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRNADVIPARRR 1149

Qy       258 GDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTM 317
             || ||:|||||||:| |||||||:||| |||||:||||||:|||||::||||||::: :
Db      1150 GDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFIPVETLDIVT 1209

Qy       318 RT 319
             ||
Db      1210 RS 1211
```

RESULT 11
GNWVJ8
genome polyprotein - hepatitis C virus (strain HC-J8)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: A40250; PQ0397; PQ0559
R;Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.;
Virology 188, 331-341, 1992
A;Title: Full-length sequence of a hepatitis C virus genome having poor homology to repo
A;Reference number: A40250; MUID:92230232; PMID:1314459
A;Accession: A40250
A;Molecule type: genomic RNA
A;Residues: 1-3033 <OKA>
A;Cross-references: GB:D10988; GB:D01221; NID:g221608; PIDN:BAA01761.1; PID:g221609
R;Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L
J. Gen. Virol. 73, 1131-1141, 1992
A;Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
A;Reference number: PQ0393; MUID:92268871; PMID:1316939
A;Accession: PQ0397
A;Molecule type: genomic RNA
A;Residues: 2678-2754 <CHA>
A;Cross-references: DDBJ:D10134
A;Experimental source: isolate E-b12
R;Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno
Biochem. Biophys. Res. Commun. 181, 279-285, 1991
A;Title: Distribution of plural HCV types in Japan.
A;Reference number: PQ0554; MUID:92068204; PMID:1720309
A;Accession: PQ0559
A;Molecule type: mRNA
A;Residues: 2678-2729 <KAT>
A;Cross-references: GB:D10562; GB:D90518; NID:g221523; PIDN:BAA01418.1; PID:g221524
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F;1011-1619/Product: hepacivirin #status predicted <NS3>
F;1234-1241/Region: nucleotide-binding motif A (P-loop)
F;1316-1321/Region: nucleotide-binding motif B
F;1320-1323/Region: DEXH motif
F;1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>
F;1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>
F;2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,23

Query Match 65.4%; Score 1158; DB 1; Length 3033;
Best Local Similarity 67.3%; Pred. No. 1.8e-89;
Matches 206; Conservative 47; Mismatches 53; Indels 0; Gaps 0;

```
Qy        16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
             | : ::|||||  |:| | ||:  ||  |:||  : :   ||||:||||:||  ||  |
Db       908 ASLLRIPYFVRAHALLRVCTLVKHLAGARYIQMLLITIGRWTGTYIYDHLSPLSTWAAQG 967

Qy        76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
             |||||:|||||:|| || |:| |||:| |||||: ||||||| |||:|||||| :  :||
Db       968 LRDLAIAVEPVVFSPMEKKVIVWGAETVACGDILHGLPVSARLGREVLLGPADGYTSKGW 1027

Qy       136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
             :|||||||:|||||||| |: |||||||:  |:|||:|: ||:|| | ::|| |||:|
Db      1028 KLLAPITAYTQQTRGLLGAIVVSLTGRDKNEQAGQVQVLSSVTQTFLGTSISGVLWTVYH 1087

Qy       196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
             |||:|||||||||:|||||: : ||||| :||| :|: ||||: |||||||:|||||||
Db      1088 GAGNKTLAGPKGPVTQMYTSAEGDLVGWPSPPGTKSLDPCTCGAVDLYLVTRNADVIPVR 1147

Qy       256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
             |: | ||:|||||||:| |||||||||:||  |||||:|||||| |||||::|||||||::
Db      1148 RKDDRRGALLSPRPLSTLKGSSGGPVLCSRGHAVGLFRAAVCARGVAKSIDFIPVESLDV 1207

Qy       316 TMRTSS 321
             || | |
Db      1208 ATRTPS 1213
```

RESULT 12
T08841
polyprotein - douroucouli hepatitis GB virus A
C;Species: douroucouli hepatitis GB virus A
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C;Accession: T08841
R;Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
J. Gen. Virol. 79, 41-45, 1998
A;Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
A;Reference number: Z16486; MUID:98120818; PMID:9460920
A;Accession: T08841
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-3005 <ERK>
A;Cross-references: EMBL:AF023425; NID:g2828599; PIDN:AAC40502.1; PID:g2828600
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: polyprotein

Query Match 22.5%; Score 398.5; DB 2; Length 3005;
Best Local Similarity 35.4%; Pred. No. 5.3e-25;
Matches 97; Conservative 45; Mismatches 125; Indels 7; Gaps 5;

```
Qy        49 AFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDI 108
             ||::    | :: |   :   | | | || ||:|||  :  :  |:  | | |||
Db       887 AFVRRLERRGVTLFQHCGQVSXGAAAILXDLGVALEPVSVTARDCYIVRDAARTLACGQR 946

Qy       109 ISGLPVSARRGREILLG--PADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQ 166
             : |||| |||| |:|:|  |:      |:    ||:    |: |    : ||: |||: :
Db       947 VEGLPVVARRGDEVLVGVFPSVRALPPGPVPTAPVVV-MQRGLGFFSVVKTSMLGRDERE 1005

Qy       167 VEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAP 226
             || : |: |:|   : |||||| :| |||: ::||||| ||:   : :   |:  :  |
Db      1006 HEGSIVVLGTSTTRSMGTCVNGVMYTTFHGSNARTLAGPVGPVNCRWWSPSDDVAVYPLP 1065

Qy       227 PGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSG 286
             ||  : || ||:  :: :    |          |  |  :|  :|||| |:||  |
Db      1066 SGASCLEPCKCGTQSVWCIRN--DGALCHGRLSKLVELDLPTEISDFRGSSGSPILCDEG 1123
```

```
Qy        287 HAVGIFRAAVCTRGVAKAVDFIPVESMETTMRTS 320
              |||:   :|  ||| |      |:  ||  : |
Db       1124 HVVGMM-VSVLHRGV-KVTGVRYVKPWETLPKDS 1155
```

RESULT 13
T08839
polyprotein - marmoset hepatitis GB virus A
C;Species: marmoset hepatitis GB virus A
C;Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 17-Nov-2000
C;Accession: T08839
R;Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
J. Gen. Virol. 79, 41-45, 1998
A;Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
A;Reference number: Z16486; MUID:98120818; PMID:9460920
A;Accession: T08839
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: genomic RNA
A;Residues: 1-2970 <ERK>
A;Cross-references: EMBL:AF023424; NID:g2828597; PIDN:AAC40501.1; PID:g2828598
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: polyprotein

  Query Match 19.4%; Score 343; DB 2; Length 2970;
  Best Local Similarity 30.5%; Pred. No. 2.7e-20;
  Matches 85; Conservative 44; Mismatches 94; Indels 56; Gaps 9;

```
Qy         71 WAHAG---------LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGRE 121
              :||||          ||    |:|||     :  ::   | | :||  :  | || |||| |
Db        891 YAHAGQVTRRTAEQLRQWGFALEPVAVHPEDCAMVRDAARTLSCGQSVHGKPVVARRGDE 950

Qy        122 ILLGPADNFEGQGWRL------LAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVS 175
              :|:|  :     | |       ||:  :    :|  | : ||:|| |: :  | | |:
Db        951 VLIGVLNGV----WELPPGFVPTAPVVVH-HHGKGFFGVVKTSMTGWDETEHVGNVVVLG 1005

Qy        176 TATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPC 235
              |:|   : |||||| :| :||: ::|||   ||:   : :   |:  :  | ||: : ||
Db       1006 TSTTRSMGTCVNGVMYTTYHGSNARTLAAQMGPVNSRWWSASDDVAVYPLPVGAKCLEPC 1065

Qy        236 TCGSSDLYLVTRHADVIPVRRRGDSRGSLLS----------PRPVSYLKGSSGGPLLCP 284
              |    ::::           | |  |:|             |  :   :|||| |:||
Db       1066 KCQPQGVWVI----------RND--GALCHGTLGRTVELDLPAELCDFRGSSGSPILCD 1112

Qy        285 SGHAVGIFRAAVCTRGVAKAVDFIPVESMETTMRTSSAW 323
              |||||:   :|  ||          |  | :| :  |
Db       1113 EGHAVGML-ISVLHRG----------SRVTGIRYTKPW 1139
```

RESULT 14
G87392
conserved hypothetical protein CC1155 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: G87392
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: G87392
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-353 <STO>
A;Cross-references: GB:AE005673; NID:g13422473; PIDN:AAK23139.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1155

  Query Match 6.1%; Score 108; DB 2; Length 353;
  Best Local Similarity 21.9%; Pred. No. 0.16;
  Matches 77; Conservative 36; Mismatches 122; Indels 116; Gaps 14;

```
Qy         22 PYFVRAQGLIRACMLVRKAA---------GGHYV-------QMAFMKLAALTGTYVYDHL 65
              |  | | ||      | |::|          |: :        :| | | || |  :
Db         65 PIAVLAGGLFAFSQLARESAIVAMRASGISGYRIVGMAVPAAVAVMLLDALCGQVLAPRA 124

Qy         66 TP-LQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILL 124
              | | ||        :|   |||  :       :  ||: :| |  |:|: :  |  : ||:
Db        125 DPTLADW-WRNTTPVAERKEPVPRTFRAGADLVIGANASADGRTITGVTIFRRDSKGILV 183

Qy        125 ----GPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQS 180
                  ||  ::|: | |  | |                :  |::   |||
Db        184 EKVEAPAARYDGKAWTLEQPKT----------------------TRFAGDLSQASTAA-- 219

Qy        181 FLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSS 240
                        |                | :     | | |  : | | |
Db        220 -------ATSWP----------------TALRPQDVQGLFGDDSMPTAAS--------- 246

Qy        241 DLYLVTRHADVIPVRRRGDSRGSLLSPRPVSY----LKGSSGGP------LLCPSGHAVG 290
                          ||  :: |   | || |:    |: :   |        ||  :  |:
Db        247 ------------ARRALENGG---SDRPESFYATHLQAAFASPFVSLVMLLLSAPVALA 290

Qy        291 IFR----AAVCTRGVAKAVDFIPVESMETTMRTSSA-------WRHPQFGG 330
              ||     | : | |:|  : |:     | | :  | |       |  |   |
Db        291 NFRSGQGAVLLTGGLAAGLMFLVANGMLTALGESGALTPFLAVWAAPAIFG 341
```

RESULT 15
B46642
DNA-directed DNA polymerase (EC 2.7.7.7) alpha/DNA primase (EC 2.7.7.-) complex 68K chair
C;Species: Mus musculus (house mouse)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Jun-2000
C;Accession: B46642
R;Miyazawa, H.; Izumi, M.; Tada, S.; Takada, R.; Masutani, M.; Ui, M.; Hanaoka, F.
J. Biol. Chem. 268, 8111-8122, 1993
A;Title: Molecular cloning of the cDNAs for the four subunits of mouse DNA polymerase alp
A;Reference number: A46642; MUID:93216788; PMID:8463324
A;Accession: B46642
A;Status: preliminary
A;Molecule type: mRNA; protein
A;Residues: 1-600 <MIY>
A;Cross-references: GB:D13546; NID:g303658; PIDN:BAA02746.1; PID:g303659
A;Experimental source: FM3A cells
A;Note: sequence extracted from NCBI backbone (NCBIN:129148, NCBIP:129149)
C;Keywords: nucleotidyltransferase

  Query Match 5.7%; Score 101; DB 2; Length 600;
  Best Local Similarity 24.8%; Pred. No. 1.3;
  Matches 55; Conservative 34; Mismatches 71; Indels 62; Gaps 12;

```
Qy         30 LIRACMLVRKAAGGHYVQM-AFMKLAALT-------GTYVYDHL---------TPLQDWA 72
              |   |:| |:   |   ::  ||   |  |         :: |: |         :  :|
Db         27 LAELCVLYRQTEDGMVSELIAFCTSAGKTCLTVDILNSFEYEVLNKKLSKAWHSASKDSG 86

Qy         73 HAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDI--ISGLP--------VSARRGREI 122
              ||| ||: |::: :| :: | : :      | : | :  :|  |         |:||  |:
Db         87 HAGTRDI-VSIQELIEAEEEEETLLSSYTTPSKGPLKRVSSTPETPLTKRSVAARSPRQ- 144

Qy        123 LLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFL 182
              || |:          :|    ||:          || | |      |||       : |
Db        145 LLSPSS---------FSPSATPSQK--------YTSRTNR------GEVVTTFGSAQ--- 178

Qy        183 ATCVNGVCWTVFHGAGSKTL--AGPKGPITQMYTNVDQDLVG 222
                |: |:   |:|| :|    |    |:|  |  : | |:|
Db        179 -----GLSWSGRGGSGSVSLKVVGDPEPLTGSYKAMFQQLMG 215
```

Search completed: May 6, 2004, 09:37:14
Job time : 12.8777 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:09:55 ; Search time 6.97286 Seconds
(without alignments)
2494.160 Million cell updates/sec

Title: US-10-650-585-4
Perfect score: 1771
Sequence: 1 MKKKKLEHHHHHHTSAGITK.....TTMRTSSAWRHPQFGGKKKK 334

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1533	86.6	3010	1	POLG_HCVJT	Q00269 h genome po
2	1528	86.3	3010	1	POLG_HCVJA	P26662 h genome po
3	1515	85.5	3010	1	POLG_HCVTW	P29846 h genome po
4	1479	83.5	3010	1	POLG_HCVBK	P26663 h genome po
5	1399	79.0	3011	1	POLG_HCV1	P26664 h genome po
6	1386	78.3	3011	1	POLG_HCVH	P27958 h genome po
7	1173	66.2	3033	1	POLG_HCVJ6	P26660 h genome po
8	1158	65.4	3033	1	POLG_HCVJ8	P26661 h genome po
9	101	5.7	600	1	DPO2_MOUSE	P33611 mus musculu
10	95.5	5.4	706	1	TRFE_HORSE	P27425 equus cabal
11	93	5.3	660	1	VST2_HEVBU	P29326 hepatitis e
12	93	5.3	660	1	VST2_HEVPA	P33426 hepatitis e
13	91.5	5.2	1705	1	PTPV_MOUSE	P70289 mus musculu
14	90	5.1	659	1	VST2_HEVME	Q03500 hepatitis e
15	88	5.0	3414	1	POLG_TBEVW	P14336 t genome po
16	87	4.9	3412	1	POLG_TBEVS	P07720 t genome po
17	87	4.9	3414	1	POLG_TBEVH	Q01299 t genome po
18	86.5	4.9	470	1	NRAM_IAWHM	P05803 influenza a
19	86	4.9	434	1	TOLB_CHLTE	Q8keq0 chlorobium
20	85	4.8	470	1	NRAM_IATRA	P03472 influenza a
21	85	4.8	730	1	HELS_METMA	Q8pzr7 methanosarc
22	84.5	4.8	347	1	MDHM_EUCGU	P46487 eucalyptus
23	84	4.7	309	1	UCP2_RAT	P56500 rattus norv
24	84	4.7	339	1	GPDA_COREF	Q8fpr0 corynebacte
25	84	4.7	470	1	NRAM_IARUE	P31510 influenza a
26	83.5	4.7	538	1	DAC_ACTSP	P39045 actinomadur
27	83.5	4.7	854	1	PWP2_SCHPO	Q9c1x1 schizosacch
28	83.5	4.7	1399	1	RPOC_PSEAE	Q9hwc9 pseudomonas
29	83	4.7	309	1	UCP2_HUMAN	P55851 homo sapien
30	83	4.7	485	1	VST2_HEVRH	Q00270 hepatitis e
31	83	4.7	660	1	VST2_HEVMY	Q04611 hepatitis e
32	82.5	4.7	453	1	NRAM_IAWIL	P03470 influenza a
33	82	4.6	309	1	UCP2_MOUSE	P70406 mus musculu
34	82	4.6	403	1	PGK_CHLMU	Q9pln4 chlamydia m
35	82	4.6	465	1	RSP4_CHLRE	Q01656 chlamydomon
36	82	4.6	612	1	AMYG_ASPOR	P36914 aspergillus
37	82	4.6	1224	1	WD11_HUMAN	Q9bzh6 homo sapien
38	81.5	4.6	398	1	TRMU_AGRT5	Q8u9m5 agrobacteri
39	81.5	4.6	467	1	GLNA_RHOSH	P43518 rhodobacter
40	81.5	4.6	1022	1	CA26_CHICK	P15988 gallus gall
41	81	4.6	209	1	PAAD_PSEAE	Q9hx08 pseudomonas
42	81	4.6	350	1	PE24_ARATH	Q9zv04 arabidopsis
43	81	4.6	470	1	NRAM_IAUSS	P03469 influenza a
44	81	4.6	511	1	VGLC_HSV11	P10228 herpes simp
45	81	4.6	730	1	HELS_METAC	Q8t139 methanosarc

ALIGNMENTS

RESULT 1
POLG_HCVJT
ID POLG_HCVJT STANDARD; PRT; 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals.";
RL Virus Res. 23:39-53(1992).
CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function. NS3
CC and NS5 may play a role in the viral RNA replication.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; D11168; BAA01943.1; -.
DR PIR; A45573; A45573.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.

```
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT   CHAIN       730   1006       NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3010       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE    1083   1083      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1107   1107      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1165   1165      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND     1230   1237      ATP (POTENTIAL).
FT   SITE        1316   1319      DECH BOX.
FT   CARBOHYD     196    196      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     209    209      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     234    234      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     250    250      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     305    305      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     417    417      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     423    423      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     430    430      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     448    448      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     532    532      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     540    540      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     556    556      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     576    576      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     623    623      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     645    645      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2041   2041      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2077   2077      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2240   2240      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2529   2529      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2788   2788      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3010 AA;  326573 MW;  94A1C77435D642BB CRC64;

  Query Match             86.6%;  Score 1533;  DB 1;  Length 3010;
  Best Local Similarity   95.1%;  Pred. No. 9.9e-124;
  Matches  289;  Conservative    7;  Mismatches    8;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              | || :|||||||||||||||||| |||||||||||||||||||||||||||||||||||
```

```
Db        904 AAITAMPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              |||||||||||:||||| |||||||||||||||| |||||||||||||||||: |||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGPADSIEGQGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              |||||||||:|||||||||:||||||||||||||||||||||||||||||||||||||||
Db       1024 RLLAPITAYAQQTRGLLGCIVTSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              |||||||||||||||||||||||||| ||||||:|||||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWHAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 RRGDGRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203

Qy        316 TMRT 319
              |||:
Db       1204 TMRS 1207
```

```
RESULT 2
POLG_HCVJA
ID   POLG_HCVJA     STANDARD;      PRT;  3010 AA.
AC   P26662;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate Japanese) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91088550; PubMed=2175903;
RA   Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA   Sugimura T., Shimotohno K.;
RT   "Molecular cloning of the human hepatitis C virus genome from
RT   Japanese patients with non-A, non-B hepatitis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RN   [2]
RP   DISCUSSION OF SEQUENCE.
RX   MEDLINE=91192160; PubMed=1849488;
RA   Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraiso K.,
RA   Ohkoshi S., Shimotohno K.;
RT   "Molecular structure of the Japanese hepatitis C viral genome.";
RL   FEBS Lett. 280:325-328(1991).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D90208; BAA14233.1; -.
DR   PIR; A39253; GNWVCJ.
DR   HSSP; P26663; 1JXP.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT   CHAIN       730   1006       NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3010       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
FT   SITE       1316   1319       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2041   2041       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2077   2077       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2240   2240       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2788   2788       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3010 AA;  327017 MW;  AA993794F46DB185 CRC64;

  Query Match             86.3%;  Score 1528;  DB 1;  Length 3010;
  Best Local Similarity   94.1%;  Pred. No. 2.7e-123;
  Matches  286;  Conservative   12;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
               ||||:|||||||||||||||||| |||||||||||||||||||||||:||||||
Db         904 AGITRVPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLRDWAHAG 963

Qy          76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
               ||||||||||:||||| |:|||||||||||||||||||:||||||||:|   |||
Db         964 LRDLAVAVEPVVFSDMETKLITWGADTAACGDIISGLPVSARRGKEILLGPADSFGEQGW 1023

Qy         136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
               ||||||||||||||||||||||||||||||:|||||:|||||||||||||||||:|
Db        1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVDGEVQVLSTATQSFLATCVNGVCWTVYH 1083

Qy         196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
               |||||||||||||||||||||||||| ||||||||||||||||||||||||:|||
Db        1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWPAPPGARSMTPCTCGSSDLYLVTRHADVVPVR 1143

Qy         256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
               |||||||||||||:|||||||||||||||| ||||||||||||||||||||||||||
Db        1144 RRGDSRGSLLSPRPISYLKGSSGGPLLCPSGHVVGIFRAAVCTRGVAKAVDFIPVESMET 1203

Qy         316 TMRT 319
               ||||:
Db        1204 TMRS 1207


RESULT 3
POLG_HCVTW
ID   POLG_HCVTW     STANDARD;      PRT;  3010 AA.
AC   P29846;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate Taiwan) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=31645;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92230206; PubMed=1314449;
RA   Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
RT   "The Taiwanese hepatitis C virus genome: sequence determination and
RT   mapping the 5' termini of viral genomic and antigenomic RNA.";
RL   Virology 188:102-113(1992).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M84754; -; NOT_ANNOTATED_CDS.
DR   PIR; A40244; GNWVTW.
DR   PDB; 1N64; 25-FEB-03.
DR   PDB; 1NS3; 08-APR-98.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CORE PROTEIN (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT   CHAIN       730   1006       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3010       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
FT   SITE       1316   1319       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    233    233       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2041   2041       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2077   2077       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2240   2240       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2529   2529       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2788   2788       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3010 AA;  327047 MW;  AAD267D55CDFE215 CRC64;

  Query Match           85.5%;  Score 1515;  DB 1;  Length 3010;
  Best Local Similarity   92.8%;  Pred. No. 3.5e-122;
  Matches  282;  Conservative   13;  Mismatches    9;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||:|::|||||||||||||||||||||||||| |||||||||||||||||||| |
Db        904 AGLTRIPYFVRAQGLIRACMLVRKAAGGHYVQMALMKLAALTGTYVYDHLTPLQDWAHTG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||:||||| |||||||||||||| ||||||||||||||||: ||:||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGPADSLEGRGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              |||||||||:|||||| ||||||||||||||||||||||||||||||||||:||||||:|
Db       1024 RLLAPITAYAQQTRGLFGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCINGVCWTVYH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              |||||||||||||||||||||||||||| || ||||:|||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWHAPQGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              |||||||||||||:|||||||||||||||| ||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPISYLKGSSGGPLLCPSGHVVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        316 TMRT 319
              ||||:
Db       1204 TMRS 1207


RESULT 4
POLG_HCVBK
ID   POLG_HCVBK     STANDARD;      PRT;  3010 AA.
AC   P26663;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate BK) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11105;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91140698; PubMed=1847440;
RA   Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RA   Onishi E., Andoh T., Yoshida I., Okayama H.;
RT   "Structure and organization of the hepatitis C virus genome isolated
RT   from human carriers.";
RL   J. Virol. 65:1105-1113(1991).
RN   [2]
RP   SEQUENCE OF 1487-1500.
RX   MEDLINE=96235224; PubMed=8647104;
RA   Borowski P., Heiland M., Oehlmann K., Becker B., Korneteky L.;
RT   "Non-structural protein 3 of hepatitis C virus inhibits
```

```
RT   phosphorylation mediated by cAMP-dependent protein kinase.";
RL   Eur. J. Biochem. 237:611-618(1996).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX   MEDLINE=97015088; PubMed=8861916;
RA   Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
RA   Moomaw E.W., Adachi T., Hostomska Z.;
RT   "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RT   trypsin-like fold and a structural zinc binding site.";
RL   Cell 87:331-342(1996).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX   MEDLINE=98227846; PubMed=9568891;
RA   Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
RA   Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
RT   "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
RT   virus: a 2.2-A resolution structure in a hexagonal crystal form.";
RL   Protein Sci. 7:837-847(1998).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M58335; AAA72945.1; -.
DR   PIR; A38465; GNWVTC.
DR   PDB; 1A1Q; 25-MAR-98.
DR   PDB; 1JXP; 14-JAN-98.
DR   PDB; 1NS3; 08-APR-98.
DR   PDB; 1C2P; 15-NOV-00.
DR   PDB; 1CSJ; 08-NOV-99.
DR   PDB; 1GX5; 09-APR-02.
DR   PDB; 1GX6; 10-APR-02.
DR   PDB; 1QUV; 26-JUN-00.
DR   PDB; 8OHM; 20-APR-99.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT   CHAIN       730   1006       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3010       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM.
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM.
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM.
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
FT   SITE       1316   1319       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2041   2041       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2077   2077       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2240   2240       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2529   2529       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2788   2788       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   STRAND     1031   1035
FT   HELIX      1039   1047
FT   STRAND     1050   1050
FT   STRAND     1059   1063
FT   STRAND     1068   1074
FT   TURN       1075   1076
FT   STRAND     1077   1081
FT   HELIX      1082   1085
FT   TURN       1086   1087
FT   STRAND     1090   1092
FT   TURN       1093   1094
FT   STRAND     1095   1097
FT   STRAND     1101   1103
FT   TURN       1104   1107
FT   STRAND     1108   1112
FT   STRAND     1120   1120
FT   STRAND     1122   1122
FT   STRAND     1129   1133
FT   TURN       1135   1136
FT   STRAND     1139   1144
FT   STRAND     1149   1157
FT   HELIX      1158   1161
FT   TURN       1162   1163
FT   TURN       1165   1166
FT   STRAND     1168   1171
FT   TURN       1172   1174
```

```
FT   STRAND      1175   1186
FT   TURN        1187   1188
FT   STRAND      1189   1197
FT   HELIX       1198   1202
FT   TURN        1203   1204
FT   STRAND      1680   1688
SQ   SEQUENCE   3010 AA;  327189 MW;  F8422D5ECCFDFD9C CRC64;

  Query Match             83.5%;  Score 1479;  DB 1;  Length 3010;
  Best Local Similarity   91.8%;  Pred. No. 4.5e-119;
  Matches  279;  Conservative   12;  Mismatches   13;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||:|||||||||||| ||||||| |||||||||||||| ||||||:|:||||:||  ||
Db        904 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWPRAG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              |||||||||||:|||||  |||||||||||||||| |||||||||:|||||||: ||:|
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGRGL 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              |||||||| |||||||||||||||||  ||||||:|||||||||||||||||||||||||
Db       1084 GAGSKTLAAPKGPITQMYTNVDQDLVGWPKPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||| |||||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPFGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        316 TMRT 319
              ||||:
Db       1204 TMRS 1207


RESULT 5
POLG_HCV1
ID   POLG_HCV1      STANDARD;      PRT;  3011 AA.
AC   P26664;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate 1) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11104;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91172826; PubMed=1848704;
RA   Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA   Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
RA   Bradley D.W., Kuo G., Houghton M.;
RT   "Genetic organization and diversity of the hepatitis C virus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M62321; AAA45676.1; -.
DR   PIR; A39166; GNWVC3.
DR   PDB; 1A1V; 16-FEB-99.
DR   PDB; 1HEI; 25-NOV-98.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT   CHAIN       730   1006       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3011       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE    1083   1083      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1107   1107      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1165   1165      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND     1230   1237      ATP (POTENTIAL).
FT   SITE        1316   1319      DECH BOX.
FT   CARBOHYD     196    196      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     209    209      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     234    234      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     305    305      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     417    417      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     423    423      N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    476    476       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2041   2041       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2077   2077       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2240   2240       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2364   2364       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2789   2789       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3011 AA;  327197 MW;  65F8C9447FCE5AF9 CRC64;

  Query Match             79.0%;  Score 1399;  DB 1;  Length 3011;
  Best Local Similarity   84.2%;  Pred. No. 3.6e-112;
  Matches  256;  Conservative   25;  Mismatches   23;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              | : |||||| |||:| | | ||  |||||||  :|| ||||||||:|||||:|||| |
Db        904 ASLLKVPYFVRVQGLLRFCALARKMIGGHYVQMVIIKLGALTGTYVYNHLTPLRDWAHNG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              |||||||||||:|| || |:|||||||||||||:|||||||||||||||     :||
Db        964 LRDLAVAVEPVVFSQMETKLITWGADTAACGDIINGLPVSARRGREILLGPADGMVSKGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              |||||||||:|||||||||||||||||||||||||||:|||| |:|||||:|||||||:|
Db       1024 RLLAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              |||::|:| |||| |||||||||||||| || |:||:|||||||||||||||||||||||
Db       1084 GAGTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              |||||||||||||:|||||||||||||:|||||||||||||||||||||||::||
Db       1144 RRGDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLET 1203

Qy        316 TMRT 319
              ||||
Db       1204 TMRS 1207


RESULT 6
POLG_HCVH
ID   POLG_HCVH      STANDARD;      PRT;  3011 AA.
AC   P27958;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate H) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11108;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92052256; PubMed=1658800;
RA   Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA   Prince A.M.;
RT   "Genomic structure of the human prototype strain H of hepatitis C
RT   virus: comparison with American and Japanese isolates.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX   MEDLINE=97331322; PubMed=9187654;
RA   Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT   "Structure of the hepatitis C virus RNA helicase domain.";
RL   Nat. Struct. Biol. 4:463-467(1997).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX   MEDLINE=98154321; PubMed=9493270;
RA   Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA   Murcko M.A., Lin C., Caron P.R.;
RT   "Hepatitis C virus NS3 RNA helicase domain with a bound
RT   oligonucleotide: the crystal structure provides insights into the mode
RT   of unwinding.";
RL   Structure 6:89-100(1998).
CC   -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC   -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC       NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC   -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC       ACTIVATION OF NS3.
CC   -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC   -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC       ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       (RNA)(N).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC       AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC   -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC       PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC   -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC   -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M67463; AAA45534.1; -.
DR   PIR; A36814; GNWVCH.
DR   PDB; 1HEI; 25-NOV-98.
DR   PDB; 1A1V; 16-FEB-99.
DR   PDB; 1A1R; 17-JUN-98.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   TRANSFAC; T04155; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
```

```
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    191       CAPSID PROTEIN C.
FT   CHAIN       192    383       ENVELOPE GLYCOPROTEIN E1.
FT   CHAIN       384    746       ENVELOPE GLYCOPROTEIN E2.
FT   CHAIN       747    809       PROTEIN P7.
FT   CHAIN       810   1026       NONSTRUCTURAL PROTEIN NS2.
FT   CHAIN      1027   1657       PROTEASE/HELICASE NS3.
FT   CHAIN      1658   1711       NONSTRUCTURAL PROTEIN NS4A.
FT   CHAIN      1712   1972       NONSTRUCTURAL PROTEIN NS4B.
FT   CHAIN      1973   2420       NONSTRUCTURAL PROTEIN NS5A.
FT   CHAIN      2421   3011       NONSTRUCTURAL PROTEIN NS5B.
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
FT   SITE       1316   1319       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    476    476       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   STRAND     1224   1226
FT   TURN       1232   1233
FT   TURN       1236   1238
FT   HELIX      1239   1246
FT   TURN       1247   1248
FT   STRAND     1251   1255
FT   HELIX      1258   1271
FT   TURN       1272   1272
FT   STRAND     1277   1280
FT   TURN       1281   1282
FT   STRAND     1283   1285
FT   STRAND     1291   1295
FT   HELIX      1296   1301
FT   TURN       1302   1303
FT   STRAND     1312   1316
FT   TURN       1317   1319
FT   HELIX      1323   1335
FT   TURN       1336   1340
FT   STRAND     1343   1347
FT   TURN       1352   1353
FT   TURN       1360   1361
FT   STRAND     1362   1366
FT   STRAND     1368   1368
FT   STRAND     1373   1375
FT   TURN       1376   1377
FT   STRAND     1378   1380
FT   HELIX      1382   1385
FT   STRAND     1389   1393
FT   HELIX      1397   1409
FT   TURN       1410   1411
FT   STRAND     1414   1417
FT   TURN       1419   1420
FT   STRAND     1432   1436
FT   TURN       1438   1439
FT   STRAND     1450   1453
FT   STRAND     1456   1463
FT   STRAND     1471   1478
FT   STRAND     1480   1480
FT   HELIX      1481   1488
FT   TURN       1489   1490
FT   STRAND     1497   1501
FT   STRAND     1507   1507
FT   STRAND     1511   1511
FT   HELIX      1514   1527
FT   HELIX      1532   1544
FT   STRAND     1550   1550
FT   HELIX      1555   1564
FT   HELIX      1570   1578
FT   TURN       1579   1580
FT   HELIX      1584   1597
FT   TURN       1598   1598
FT   HELIX      1606   1611
FT   TURN       1614   1618
FT   STRAND     1622   1623
FT   STRAND     1627   1627
FT   STRAND     1635   1636
FT   HELIX      1640   1652
SQ   SEQUENCE   3011 AA;  327142 MW;  772CBB29CCD94753 CRC64;

  Query Match           78.3%;  Score 1386;  DB 1;  Length 3011;
  Best Local Similarity  83.6%;  Pred. No. 4.8e-111;
  Matches  254;  Conservative   27;  Mismatches   23;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              | : ||||||| |||:| | | || ||||||||| :|| ||||| ||:|| ||:|||| |
Db        904 ASLLKVPYFVRVQGLLRICALARKIAGGHYVQMAIIKLGALTGTCVYNHLAPLRDWAHNG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||:|| || |:|||||||||||||:|||||||:||||||||    :||
Db        964 LRDLAVAVEPVVFSRMETKLITWGADTAACGDIINGLPVSARRGQEILLGPADGMVSKGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              |||||||||:|||||||||||||||||||||||||||:||||||:|||||:||||||:|
Db       1024 RLLAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTATQTFLATCINGVCWTVYH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              |||::|:| |||| | |||||||||| || |:||:|||||||||||||||||||||||
Db       1084 GAGTRTIASPKGPVIQTYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||:|||||||||||||:|||||:|||||||||||||||||||::||
Db       1144 RRGDSRGSLLSPRPISYLKGSSGGPLLCPTGHAVGLFRAAVCTRGVAKAVDFIPVENLET 1203

Qy        316 TMRT 319
              |||:
Db       1204 TMRS 1207


RESULT 7
POLG_HCVJ6
ID   POLG_HCVJ6     STANDARD;      PRT;  3033 AA.
AC   P26660;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate HC-J6) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
```

```
OC   Hepacivirus.
OX   NCBI_TaxID=11113;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92044440; PubMed=1658196;
RA   Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
RA   Machida A., Miyakawa Y., Mayumi M.;
RT   "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
RT   from a human carrier: comparison with reported isolates for conserved
RT   and divergent regions.";
RL   J. Gen. Virol. 72:2697-2704(1991).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       (RNA)(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D00944; BAA00792.1; -.
DR   PIR; JQ1303; JQ1303.
DR   HSSP; P27958; 1HEI.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    733       NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT   CHAIN       734   1010       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1011   1619       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1620   1866       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1867   2017       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2018   3033       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE    1087   1087      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1111   1111      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1169   1169      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND     1234   1241      ATP (POTENTIAL).
FT   SITE        1320   1323      DECH BOX.
FT   CARBOHYD     196    196      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     209    209      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     234    234      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     305    305      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     417    417      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     423    423      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     430    430      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     448    448      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     477    477      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     534    534      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     542    542      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     558    558      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     578    578      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     627    627      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     649    649      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1091   1091      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2038   2038      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2811   2811      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3033 AA;  329165 MW;  F957F5C1A273BE9E CRC64;

  Query Match             66.2%;  Score 1173;  DB 1;  Length 3033;
  Best Local Similarity   69.2%;  Pred. No. 1.2e-92;
  Matches  209;  Conservative   45;  Mismatches   48;  Indels    0;  Gaps    0;

Qy         18 ITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAGLR 77
              :|:|||||||  |:| | :||  ||| ||||  : |   ||||:||||||: |||  |||
Db        910 LTRVPYFVRAHALLRMCTMVRHLAGGRYVQMVLLALGRWTGTYIYDHLTPMSDWAANGLR 969

Qy         78 DLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGWRL 137
              ||||||||:||| || |:| |||:|||||||: ||||||| |||:|||||| :  :|| |
Db        970 DLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPADGYTSKGWSL 1029

Qy        138 LAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFHGA 197
              |||||||:|||||||| |: |:||||| :  ||:||:|| |||||  | ::|| |||:|||
Db       1030 LAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTTISGVLWTVYHGA 1089

Qy        198 GSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVRRR 257
              |:||||| :||:||||:: : ||||| :||| :|: |||||: |||||||:||||| |||
Db       1090 GNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRNADVIPARRR 1149

Qy        258 GDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTM 317
              || ||:||||||:| |||||||:||| |||||:||||||:|||||::||||||:::
Db       1150 GDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFIPVETLDIVT 1209

Qy        318 RT 319
              |:
Db       1210 RS 1211


RESULT 8
POLG_HCVJ8
ID   POLG_HCVJ8     STANDARD;      PRT;  3033 AA.
AC   P26661;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
```

```
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate HC-J8) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11115;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92230232; PubMed=1314459;
RA   Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
RA   Fukuda S., Tsuda F., Mishiro S.;
RT   "Full-length sequence of a hepatitis C virus genome having poor
RT   homology to reported isolates: comparative study of four distinct
RT   genotypes.";
RL   Virology 188:331-341(1992).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10988; BAA01761.1; -.
DR   PIR; A40250; GNWVJ8.
DR   HSSP; P27958; 1HEI.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    733       NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT   CHAIN       734   1010       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1011   1619       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1620   1866       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1867   2017       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2018   3033       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE    1087   1087      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1111   1111      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1169   1169      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND     1234   1241      ATP (POTENTIAL).
FT   SITE        1320   1323      DECH BOX.
FT   CARBOHYD     196    196      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     209    209      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     233    233      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     299    299      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     305    305      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     417    417      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     423    423      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     430    430      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     448    448      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     477    477      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     534    534      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     542    542      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     558    558      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     578    578      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     627    627      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     649    649      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1091   1091      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2038   2038      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2359   2359      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2811   2811      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3033 AA;  330177 MW;  1A173E7E3381FD1A CRC64;

  Query Match             65.4%;  Score 1158;  DB 1;  Length 3033;
  Best Local Similarity   67.3%;  Pred. No. 2.3e-91;
  Matches  206;  Conservative   47;  Mismatches   53;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              | : ::|||||  |:| | ||:  ||  |:||  : :   ||||:||||:||  ||  |
Db        908 ASLLRIPYFVRAHALLRVCTLVKHLAGARYIQMLLITIGRWTGTYIYDHLSPLSTWAAQG 967

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              |||||:|||||:|| || |:| |||:| |||||: |||||| |||:|||||| :  :||
Db        968 LRDLAIAVEPVVFSPMEKKVIVWGAETVACGDILHGLPVSARLGREVLLGPADGYTSKGW 1027

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              :||||||||:||||||| |: ||||||||:  |:|||:|: ||:|| | ::|| |||:|
Db       1028 KLLAPITAYTQQTRGLLGAIVVSLTGRDKNEQAGQVQVLSSVTQTFLGTSISGVLWTVYH 1087

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              |||:|||||||||:|||||: : ||||| :||| :|: |||||: |||||||:|||||||
Db       1088 GAGNKTLAGPKGPVTQMYTSAEGDLVGWPSPPGTKSLDPCTCGAVDLYLVTRNADVIPVR 1147

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              |: | ||:||||||:| ||||||||:||  |||||:|||||| |||||::|||||||::
Db       1148 RKDDRRGALLSPRPLSTLKGSSGGPVLCSRGHAVGLFRAAVCARGVAKSIDFIPVESLDV 1207

Qy        316 TMRTSS 321
              || |
Db       1208 ATRTPS 1213


RESULT 9
DPO2_MOUSE
ID   DPO2_MOUSE     STANDARD;      PRT;   600 AA.
AC   P33611;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
```

```
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   DNA polymerase alpha 70 kDa subunit (DNA polymerase subunit B).
GN   POLA2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 84-102; 269-285 AND 394-403.
RX   MEDLINE=93216788; PubMed=8463324;
RA   Miyazawa H., Izumi M., Tada S., Takada R., Masutani M., Ui M.,
RA   Hanaoka F.;
RT   "Molecular cloning of the cDNAs for the four subunits of mouse DNA
RT   polymerase alpha-primase complex and their gene expression during
RT   cell proliferation and the cell cycle.";
RL   J. Biol. Chem. 268:8111-8122(1993).
CC   -!- FUNCTION: May play an essential role at the early stage of
CC       chromosomal DNA replication by coupling the polymerase
CC       alpha/primase complex to the cellular replication machinery (By
CC       similarity).
CC   -!- SUBUNIT: DNA polymerase alpha-primase is a four subunit enzyme
CC       (subunits A, B, C and D), which is assembled throughout the cell
CC       cycle. The largest subunit (subunit A) has DNA polymerase
CC       activity, the two smallest subunits (subunits C and D) have DNA
CC       primase activity. Subunit B binds to subunit A.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- PTM: PHOSPHORYLATED IN A CELL CYCLE-DEPENDENT MANNER, IN G2/M
CC       PHASE (BY SIMILARITY).
CC   -!- SIMILARITY: Belongs to the DNA polymerase alpha subunit B family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D13546; BAA02746.1; -.
DR   PIR; B46642; B46642.
DR   MGD; MGI:99690; Pola2.
DR   InterPro; IPR007200; DNA_pol_alpha_B.
DR   Pfam; PF04058; DNA_pol_alpha_B; 1.
KW   DNA replication; Nuclear protein; Phosphorylation.
FT   DOMAIN      101    107       POLY-GLU.
FT   DOMAIN      115    157       PRO/SER/THR-RICH (HYDROPHILIC).
SQ   SEQUENCE   600 AA;  66267 MW;  79F94BE6EF33FEBC CRC64;

  Query Match             5.7%;  Score 101;  DB 1;  Length 600;
  Best Local Similarity   24.8%;  Pred. No. 0.52;
  Matches   55;  Conservative   34;  Mismatches   71;  Indels   62;  Gaps   12;

Qy         30 LIRACMLVRKAAGGHYVQM-AFMKLAALT-------GTYVYDHL---------TPLQDWA 72
              |   |:| |:   |   :: ||  | |          :: |: |         : :|
Db         27 LAELCVLYRQTEDGMVSELIAFCTSAGKTCLTVDILNSFEYEVLNKKLSKAWHSASKDSG 86

Qy         73 HAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDI--ISGLP--------VSARRGREI 122
              ||| ||: |::: :| :: | : :     | : | :  :|  |        |:|| |:
Db         87 HAGTRDI-VSIQELIEAEEEEETLLSSYTTPSKGPLKRVSSTPETPLTKRSVAARSPRQ- 144

Qy        123 LLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFL 182
              || |:           :|    ||:          || | |       |||     : |
Db        145 LLSPSS---------FSPSATPSQK--------YTSRTNR------GEVVTTFGSAQ--- 178

Qy        183 ATCVNGVCWTVFHGAGSKTL--AGPKGPITQMYTNVDQDLVG 222
                 |: |:   |:|| :|    |   |:|  |  : | |:|
Db        179 -----GLSWSGRGGSGSVSLKVVGDPEPLTGSYKAMFQQLMG 215


RESULT 10
TRFE_HORSE
ID   TRFE_HORSE     STANDARD;      PRT;   706 AA.
AC   P27425;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Serotransferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal
DE   binding globulin).
GN   TF.
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93277958; PubMed=8504171;
RA   Carpenter M.A., Broad T.E.;
RT   "The cDNA sequence of horse transferrin.";
RL   Biochim. Biophys. Acta 1173:230-232(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Extraembryonic tissue;
RA   McDowell K.J., Adams M.H., Baker C.B.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Transferrins are iron binding transport proteins which
CC       can bind two atoms of ferric iron in association with the binding
CC       of an anion, usually bicarbonate. It is responsible for the
CC       transport of iron from sites of absorption and heme degradation to
CC       those of storage and utilization. Serum transferrin may also have
CC       a further role in stimulating cell proliferation.
CC   -!- SUBUNIT: Monomer.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.
CC   -!- DOMAIN: Composed of two homologous domains.
CC   -!- SIMILARITY: Belongs to the transferrin family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M69020; AAA30958.1; -.
DR   EMBL; U21127; AAA63684.1; -.
DR   PIR; S33761; S33761.
DR   HSSP; P02787; 1A8E.
DR   InterPro; IPR001156; Transferrin.
DR   Pfam; PF00405; transferrin; 2.
DR   PRINTS; PR00422; TRANSFERRIN.
DR   SMART; SM00094; TR_FER; 2.
DR   PROSITE; PS00205; TRANSFERRIN_1; 2.
DR   PROSITE; PS00206; TRANSFERRIN_2; 2.
DR   PROSITE; PS00207; TRANSFERRIN_3; 2.
KW   Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
KW   Signal.
FT   SIGNAL        1     19       BY SIMILARITY.
FT   CHAIN        20    706       SEROTRANSFERRIN.
FT   REPEAT       20    357       1.
FT   REPEAT      358    706       2.
FT   DISULFID     26     64       BY SIMILARITY.
FT   DISULFID     36     55       BY SIMILARITY.
FT   DISULFID    134    215       BY SIMILARITY.
FT   DISULFID    174    190       BY SIMILARITY.
FT   DISULFID    177    198       BY SIMILARITY.
FT   DISULFID    187    200       BY SIMILARITY.
FT   DISULFID    248    262       BY SIMILARITY.
FT   DISULFID    360    623       BY SIMILARITY.
FT   DISULFID    366    398       BY SIMILARITY.
FT   DISULFID    376    389       BY SIMILARITY.
FT   DISULFID    423    701       BY SIMILARITY.
FT   DISULFID    441    664       BY SIMILARITY.
FT   DISULFID    474    550       BY SIMILARITY.
FT   DISULFID    498    692       BY SIMILARITY.
FT   DISULFID    508    522       BY SIMILARITY.
```

```
FT   DISULFID    519    533       BY SIMILARITY.
FT   DISULFID    590    604       BY SIMILARITY.
FT   DISULFID    642    647       BY SIMILARITY.
FT   METAL        79     79       IRON 1 (BY SIMILARITY).
FT   METAL       111    111       IRON 1 (BY SIMILARITY).
FT   METAL       209    209       IRON 1 (BY SIMILARITY).
FT   METAL       270    270       IRON 1 (BY SIMILARITY).
FT   METAL       413    413       IRON 2 (BY SIMILARITY).
FT   METAL       449    449       IRON 2 (BY SIMILARITY).
FT   METAL       544    544       IRON 2 (BY SIMILARITY).
FT   METAL       612    612       IRON 2 (BY SIMILARITY).
FT   BINDING     136    136       CARBONATE 1 (BY SIMILARITY).
FT   BINDING     140    140       CARBONATE 1 (BY SIMILARITY).
FT   BINDING     142    142       CARBONATE 1 (VIA AMIDE NITROGEN) (BY
FT                                SIMILARITY).
FT   BINDING     143    143       CARBONATE 1 (VIA AMIDE NITROGEN) (BY
FT                                SIMILARITY).
FT   BINDING     476    476       CARBONATE 2 (BY SIMILARITY).
FT   BINDING     480    480       CARBONATE 2 (BY SIMILARITY).
FT   BINDING     482    482       CARBONATE 2 (VIA AMIDE NITROGEN) (BY
FT                                SIMILARITY).
FT   BINDING     483    483       CARBONATE 2 (VIA AMIDE NITROGEN) (BY
FT                                SIMILARITY).
FT   CARBOHYD    515    515       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   706 AA;  78094 MW;  1A0FA566C0409D8A CRC64;

  Query Match             5.4%;  Score 95.5;  DB 1;  Length 706;
  Best Local Similarity   21.8%;  Pred. No. 1.9;
  Matches   67;  Conservative   43;  Mismatches  115;  Indels   83;  Gaps   18;

Qy         48 MAFMKLAALTGTYV---YDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGA---- 100
              : |:::     |::   |:::|         :|:|   : | :  | | | : | |
Db        321 LGFLRIPPAMDTWLYLGYEYVT--------AIRNLREDIRPEVPKD-ECKKVKWCAIGHH 371

Qy        101 DTAACGD--IISGLPVSARRGR-------EILLGPADNFEGQGWRL-------LAPITAY 144
              :   | :  : || :     :       :|: | ||      |  :       | |: |
Db        372 EKVKCDEWSVNSGGNIECESAQSTEDCIAKIVKGEADAMSLDGGFIYIAGKCGLVPVLAE 431

Qy        145 SQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAG 204
              : :||    |: |   |         | || :::   |         |         :| |
Db        432 NYETRSGSACVDTPEEGYH------AVAVVKSSSDPDLT-------W--------NSLKG 470

Qy        205 PKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSL 264
               |      :| ||:   ||  | |             ||   :| :     | | : |
Db        471 KK----SCHTGVDR-TAGWNIPMGL-----------LYSEIKHCEFDKFFREGCAPGYR 513

Qy        265 LSPRPVSYLKGSSGGP-LLC-PSGHA-----VGIFRAAVCTRGVAKAVDFIPVESME--T 315
               :   :  ||: ||   | |: |       | ||  |    ||    |:  :::|  |
Db        514 RNSTLCNLCIGSASGPGRECEPNNHERYYGYTGAFRCLVEKGDVA----FVKHQTVEQNT 569

Qy        316 TMRTSSAW 323
              |   |
Db        570 DGRNPDDW 577


RESULT 11
VST2_HEVBU
ID   VST2_HEVBU     STANDARD;      PRT;   660 AA.
AC   P29326;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   Structural protein 2 precursor (ORF2).
OS   Hepatitis E virus (strain Burma) (HEV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage;
OC   Hepatitis E-like viruses.
OX   NCBI_TaxID=31767;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92024067; PubMed=1926770;
RA   Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,
RA   Fry K.E., Reyes G.R.;
RT   "Hepatitis E virus (HEV): molecular cloning and sequencing of the
RT   full-length viral genome.";
RL   Virology 185:120-131(1991).
CC   -!- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
CC       THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
CC       BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M73218; AAA45736.1; -.
DR   PIR; C40778; VHWWH2.
DR   InterPro; IPR004261; SP2.
DR   InterPro; IPR008975; Viral_cap_coat.
DR   Pfam; PF03014; SP2; 1.
KW   Signal.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    660       STRUCTURAL PROTEIN 2.
SQ   SEQUENCE   660 AA;  70978 MW;  5832A013CCC4A61C CRC64;

  Query Match             5.3%;  Score 93;  DB 1;  Length 660;
  Best Local Similarity   19.0%;  Pred. No. 2.8;
  Matches   72;  Conservative   47;  Mismatches  119;  Indels  140;  Gaps   16;

Qy         27 AQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPV 86
              |:  ||   ||  | ||: : ::|     | | |        |       |: : |:|
Db        188 ARATIRYRPLVPNAVGGYAISISFWPQTTTTPTSV--------DMNSITSTDVRILVQPG 239

Qy         87 IFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPAD--NFEGQGWRLLAPI-TA 143
              | |:: :                                  |::  ::  |||| :     |
Db        240 IASELVI--------------------------------PSERLHYRNQGWRSVETSGVA 267

Qy        144 YSQQTRGLL-----GCIITSLTG----------------------------------- 161
              : | ||:     | :: | |
Db        268 EEEATSGLVMLCIHGSLVNSYTNTPYTGALGLLDFALELEFRNLTPGNTNTRVSRYSSTA 327

Qy        162 --RDKNQVEGEVQVVSTATQSFLA----TCVNGV-------CWTVFH------------ 195
                | :   :|  :: :||    |:      |  |||          |:|:
Db        328 RHRLRRGADGTAELTTTAATRFMKDLYFTSTNGVGEIGRGIALTLFNLADTLLGGLPTEL 387

Qy        196 --GAG-----SKTLAGPKG-PITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLV-- 245
                ||      |: :    | |  ::||:|:      |    |          | | : :
Db        388 ISSAGGQLFYSRPVVSANGEPTVKLYTSVENA----QQDKGIAIPHDIDLGESRVVIQDY 443

Qy        246 -TRHADVIPVRRRGDSRG-SLLSPRPVSYLK-----------GSSGGPLLCPSGHAVGIF 292
               :|    |      ||  |:|    | :|              ||| ||:       :| :
Db        444 DNQHEQDRPTPSPAPSRPFSVLRANDVLWLSLTAAEYDQSTYGSSTGPVYV--SDSVTLV 501

Qy        293 RAAVCTRGVAKAVDFIPV 310
               |    : ||:::|:  |
Db        502 NVATGAQAVARSLDWTKV 519


RESULT 12
VST2_HEVPA
ID   VST2_HEVPA     STANDARD;      PRT;   660 AA.
AC   P33426;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   Structural protein 2 precursor (ORF2).
OS   Hepatitis E virus (strain Pakistan) (HEV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage;
OC   Hepatitis E-like viruses.
OX   NCBI_TaxID=33774;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92115700; PubMed=1731327;
RA   Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,
```

```
RA   Malik I.A., Iqbal M., Purcell R.H.;
RT   "Characterization of a prototype strain of hepatitis E virus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).
CC   -!- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
CC       THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
CC       BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M80581; AAA45727.1; -.
DR   InterPro; IPR004261; SP2.
DR   InterPro; IPR008975; Viral_cap_coat.
DR   Pfam; PF03014; SP2; 1.
KW   Signal.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   CHAIN        23    660       STRUCTURAL PROTEIN 2.
SQ   SEQUENCE   660 AA;  70980 MW;  8085BC53CFB46FD3 CRC64;

  Query Match             5.3%;  Score 93;  DB 1;  Length 660;
  Best Local Similarity   19.0%;  Pred. No. 2.8;
  Matches   72;  Conservative   49;  Mismatches  117;  Indels  140;  Gaps   17;

Qy          27 AQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPV 86
               |:  ||   ||  | ||:: : ::|      | | |       |       |: : |:|
Db         188 ARATIRYRPLVPNAVGGYAISISFWPQTTTTPTSV--------DMNSITSTDVRILVQPG 239

Qy          87 IFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPAD--NFEGQGWRLL------ 138
               | |:: :                                  |::  ::  |||| :
Db         240 IASELVI--------------------------------PSERLHYRNQGWRSVETSGVA 267

Qy         139 ---------------APITAYSQQT-RGLLGCI-------ITSLTGRDKNQ--------- 166
                            :|: :|:     | || :         :||  : |
Db         268 EEEATSGLVMLCIHGSPVNSYTNTPYTGALGLLDFALELEFRNLTPGNTNTRVSRYSSTA 327

Qy         167 -------VEGEVQVVSTATQSFLA----TCVNGV-------CWTVFH------------- 195
                     :|  :: :||   |:     |  |||         |:|:
Db         328 RHRLRRGADGTAELTTTAATRFMKDLYFTSTNGVGEIGRGIALTLFNLADTLLGGLPTEL 387

Qy         196 --GAG-----SKTLAGPKG-PITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLV-- 245
                 ||     |: :   | |  ::||:|:      |   |       | | : :
Db         388 ISSAGGQLFYSRPVVSANGEPTVKLYTSVENA----QQDKGIAIPHDIDLGESRVVIQDY 443

Qy         246 -TRHADVIPVRRRGDSRG-SLLSPRPVSYLK-----------GSSGGPLLCPSGHAVGIF 292
                :|    |       ||  |:|    | :|           ||| ||:      :| :
Db         444 DNQHEQDRPTPSPAPSRPFSVLRANDVLWLSLTAAEYDQSTYGSSTGPVYV--SDSVTLV 501

Qy         293 RAAVCTRGVAKAVDFIPV 310
                 |   : ||:::|:  |
Db         502 NVATGAQAVARSLDWTKV 519


RESULT 13
PTPV_MOUSE
ID   PTPV_MOUSE     STANDARD;      PRT;  1705 AA.
AC   P70289;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Receptor-type protein-tyrosine phosphatase V precursor (EC 3.1.3.48)
DE   (Embryonic stem cell protein-tyrosine phosphatase) (ES cell
DE   phosphatase).
GN   PTPRV OR ESP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryonic stem cells;
RX   MEDLINE=97109513; PubMed=8951793;
RA   Lee K., Nichols J., Smith A.;
RT   "Identification of a developmentally regulated protein tyrosine
RT   phosphatase in embryonic stem cells that is a marker of
RT   pluripotential epiblast and early mesoderm.";
RL   Mech. Dev. 59:153-164(1996).
RN   [2]
RP   ERRATUM.
RA   Lee K., Nichols J., Smith A.;
RL   Mech. Dev. 61:213-215(1996).
CC   -!- FUNCTION: May play a role in the maintenance of pluripotency.
CC       Down-regulated during differentiation.
CC   -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC       tyrosine + phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- DEVELOPMENTAL STAGE: Detectable in the epiblast of oocytes and
CC       throughout early mouse embryo development. In adult, expression is
CC       localized in gonadal germ cells.
CC   -!- SIMILARITY: Contains 2 protein-tyrosine phosphatase domains.
CC   -!- SIMILARITY: Contains 10 fibronectin type III domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U36488; AAC52868.1; -.
DR   HSSP; P18052; 1YFO.
DR   MGD; MGI:108027; Ptprv.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR000387; TYR_phosphatase.
DR   InterPro; IPR000242; Tyr_PP.
DR   Pfam; PF00041; fn3; 7.
DR   Pfam; PF00102; Y_phosphatase; 1.
DR   PRINTS; PR00700; PRTYPHPHTASE.
DR   SMART; SM00060; FN3; 8.
DR   SMART; SM00194; PTPc; 1.
DR   PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR   PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR   PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 2.
KW   Hydrolase; Transmembrane; Repeat; Signal; Glycoprotein.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19   1705       RECEPTOR-TYPE PROTEIN-TYROSINE
FT                                PHOSPHATASE V.
FT   DOMAIN       19   1077       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1078   1100       POTENTIAL.
FT   DOMAIN     1101   1705       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       35    110       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      126    199       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      216    288       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      304    373       FIBRONECTIN TYPE-III 4.
FT   DOMAIN      393    454       FIBRONECTIN TYPE-III 5.
FT   DOMAIN      471    543       FIBRONECTIN TYPE-III 6.
FT   DOMAIN      563    634       FIBRONECTIN TYPE-III 7.
FT   DOMAIN      657    722       FIBRONECTIN TYPE-III 8.
FT   DOMAIN      742    813       FIBRONECTIN TYPE-III 9.
FT   DOMAIN      831    905       FIBRONECTIN TYPE-III 10.
FT   DOMAIN     1150   1418       PROTEIN-TYROSINE PHOSPHATASE 1.
FT   DOMAIN     1469   1700       PROTEIN-TYROSINE PHOSPHATASE 2.
FT   ACT_SITE   1350   1350       PHOSPHOCYSTEINE INTERMEDIATE (BY
FT                                SIMILARITY).
FT   CARBOHYD     74     74       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     89     89       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    117    117       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    239    239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    259    259       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    299    299       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    345    345       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    431    431       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    551    551       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    570    570       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    620    620       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    649    649       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    663    663       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    737    737       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    851    851       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    882    882       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    970    970       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    982    982       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1705 AA;  186795 MW;  2783755F15387D5B CRC64;

  Query Match             5.2%;  Score 91.5;  DB 1;  Length 1705;
  Best Local Similarity   25.7%;  Pred. No. 12;
  Matches   83;  Conservative   21;  Mismatches  120;  Indels   99;  Gaps   21;

Qy         65 LTPLQD--WAH----AGLRD-----LAVAVEPVIFSDMEVKIITWGADTAACGDIISG-- 111
              |||| :  ||       | ||     |: ||              | | :   |  : |
Db        314 LTPLWNELWASWKAGQGARDGYVLKLSGPVE---------NTTTLGPE--ECNAVFPGPL 362

Qy        112 LPVSARRGREILLGPADNF-EGQGWRLLAPITAYSQQTRG----LLGCIITSLTGR---- 162
               |     |  :| || | : ||  | .||    |   :  |    | |   |  ||
Db        363 PPGHYTLGLRVLAGPYDAWVEGSIW--LAESAARPMEVPGARLWLEGLEATKQPGRRALL 420

Qy        163 ---DKNQVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYT----- 214
                 |   : | : | | ||       | |  : |        :|   | |||  |
Db        421 YSVDAPGLLGNISVSSGATHVTFCGLVPGAHYRV-------DIASSMGDITQSLTGYTSP 473

Qy        215 -----------NVDQDL-VGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSR- 261
                        |   || :||    |           |  : | || | |        | |
Db        474 LPPQSLEIISRNSPSDLTIGWAPAP----------GQMEGYKVTWHQD-------GSQRS 516

Qy        262 -GSLLSPRP-VSYLKGSSGGPLLCPSGHAVGIF-----------RAAVCTRGVAKAVDFI 308
               | |:   | :| |   |  | |  |   : |  :           :   |||       :
Db        517 PGDLVDLGPDISSLTLKSLVPGSC---YTVSAWAWSGNLSSDSQKIHSCTRPAPPTNLSL 573

Qy        309 PVESMETTMRTSSAWRHPQFGGK 331
                 |:| |  | ||  ||:
Db        574 GFAHQPATLRAS--WCHPP-GGR 593


RESULT 14
VST2_HEVME
ID   VST2_HEVME     STANDARD;      PRT;   659 AA.
AC   Q03500;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   Structural protein 2 precursor.
OS   Hepatitis E virus (strain Mexico) (HEV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage;
OC   Hepatitis E-like viruses.
OX   NCBI_TaxID=31768;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93079857; PubMed=1448913;
RA   Huang C.C., Nguyen D., Fernandez J., Yun K.Y., Fry K.E.,
RA   Bradley D.W., Tam A.W., Reyes G.R.;
RT   "Molecular cloning and sequencing of the Mexico isolate of hepatitis
RT   E virus (HEV).";
RL   Virology 191:550-558(1992).
CC   -!- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
CC       THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
CC       BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M74506; AAA45732.1; -.
DR   PIR; B44212; B44212.
DR   InterPro; IPR004261; SP2.
DR   InterPro; IPR008975; Viral_cap_coat.
DR   Pfam; PF03014; SP2; 1.
KW   Signal.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   CHAIN        23    659       STRUCTURAL PROTEIN 2.
SQ   SEQUENCE   659 AA;  70640 MW;  CF75E75EFD8FBE2C CRC64;

  Query Match             5.1%;  Score 90;  DB 1;  Length 659;
  Best Local Similarity   17.6%;  Pred. No. 5.2;
  Matches   69;  Conservative   54;  Mismatches  124;  Indels  146;  Gaps   15;

Qy         27 AQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPV 86
              |:  ||   ||  | ||: : ::|      | | |         |       |: : |:|
Db        188 ARATIRYRPLVPNAVGGYAISISFWPQTTTTPTSV--------DMNSITSTDVRILVQPG 239

Qy         87 IFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPAD--NFEGQGWRLL------ 138
              | |:: :                                   |::  ::  |||| :
Db        240 IASELVI--------------------------------PSERLHYRNQGWRSVETSGVA 267

Qy        139 --------------APITAYSQ---------------QTRGLLGCIITSLTGR----- 162
                           :|: :|:               : | |  |   :   |
Db        268 EEEATSGLVMLCIHGSPVNSYTNTPYTGALGLLDFALELEFRNLTTCNTNTRVSRYSSTA 327

Qy        163 --DKNQVEGEVQVVSTATQSFLA----TCVNGV-------------CWTVFHGAGSKTL 202
                  :|  :: :||   |:     | :|||                |:  |  :: :
Db        328 RHSARGADGTAELTTTAATRFMKDLHFTGLNGVGEVGRGIALTLLNLADTLLGGLPTELI 387

Qy        203 AGPKG--------------PITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLV--- 245
              :  |              |  ::||:|:      |   |         | | : :
Db        388 SSAGGQLFYSRPVVSANGEPTVKLYTSVENA----QQDKGVAIPHDIDLGDSRVVIQDYD 443

Qy        246 TRHADVIPVRRRGDSRG-SLLSPRPVSYLK-----------GSSGGPLLCPSGHAVGIFR 293
               :|    |          ||  |:|    | :|           ||| ||:      :| :
Db        444 NQHEQDRPTPSPAPSRPFSVLRANDVLWLSLTAAEYDQSTYGSSTGPVYI--SDSVTLVN 501

Qy        294 AAVCTRGVAKAVDFI--------PVESMETTMRT 319
               |   : ||:::|:         |: ::|   :|
Db        502 VATGAQAVARSLDWSKVTLDGRPLPTVEQYSKT 534


RESULT 15
POLG_TBEVW
ID   POLG_TBEVW     STANDARD;      PRT;  3414 AA.
AC   P14336; Q88493;
DT   01-JAN-1990 (Rel. 13, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
DE   protein (Envelope protein M); Major envelope protein E; Nonstructural
DE   proteins NS1, NS2A, NS2B, NS4A and NS4B; Protease/helicase
DE   (EC 3.4.21.98) (NS3); RNA-directed RNA polymerase (EC 2.7.7.48)
DE   (NS5)].
OS   Tick-borne encephalitis virus (Western subtype) (TBEV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Flavivirus.
OX   NCBI_TaxID=11088;
RN   [1]
RP   SEQUENCE FROM N.A., AND REVISIONS.
RC   STRAIN=Neudoerfl;
RX   MEDLINE=96036491; PubMed=7483260;
RA   Wallner G., Mandl C.W., Kunz C., Heinz F.X.;
RT   "The flavivirus 3'-noncoding region: extensive size heterogeneity
RT   independent of evolutionary relationships among strains of tick-borne
RT   encephalitis virus.";
RL   Virology 213:169-178(1995).
RN   [2]
RP   SEQUENCE OF 1-779 FROM N.A.
```

```
RC   STRAIN=Neudoerfl;
RX   MEDLINE=88322870; PubMed=3413985;
RA   Mandl C.W., Heinz F.X., Kunz C.;
RT   "Sequence of the structural proteins of tick-borne encephalitis virus
RT   (western subtype) and comparative analysis with other flaviviruses.";
RL   Virology 166:197-205(1988).
RN   [3]
RP   SEQUENCE OF 767-3414 FROM N.A.
RC   STRAIN=Neudoerfl;
RX   MEDLINE=90051080; PubMed=2554575;
RA   Mandl C.W., Heinz F.X., Stoeckl E., Kunz C.;
RT   "Genome sequence of tick-borne encephalitis virus (Western subtype)
RT   and comparative analysis of nonstructural proteins with other
RT   flaviviruses.";
RL   Virology 173:291-301(1989).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 281-680.
RX   MEDLINE=95272700; PubMed=7753193;
RA   Rey F.A., Heinz F.X., Mandl C.W., Kunz C., Harrison S.C.;
RT   "The envelope glycoprotein from tick-borne encephalitis virus at 2-A
RT   resolution.";
RL   Nature 375:291-298(1995).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function.
CC       NS3 and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U27495; AAA86870.1; -.
DR   PIR; A31052; GNWVNE.
DR   PDB; 1SVB; 10-JUN-96.
DR   MEROPS; S07.UPW; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001122; Flavi_capsidC.
DR   InterPro; IPR000336; Flavi_glycoprotE.
DR   InterPro; IPR000069; Flavi_M.
DR   InterPro; IPR001157; Flavi_NS1.
DR   InterPro; IPR000752; Flavi_NS2A.
DR   InterPro; IPR000487; Flavi_NS2B.
DR   InterPro; IPR000404; Flavi_NS4A.
DR   InterPro; IPR001528; Flavi_NS4B.
DR   InterPro; IPR000208; Flavi_NS5.
DR   InterPro; IPR002535; Flavi_propep.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001850; Peptidase_S7.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   InterPro; IPR002877; RrmJ_FtsJ.
DR   Pfam; PF01003; Flavi_capsid; 1.
DR   Pfam; PF02832; Flavi_glycop_C; 1.
DR   Pfam; PF00869; Flavi_glycoprot; 1.
DR   Pfam; PF00949; Flavi_helicase; 1.
DR   Pfam; PF01004; Flavi_M; 1.
DR   Pfam; PF00948; Flavi_NS1; 1.
DR   Pfam; PF01005; Flavi_NS2A; 1.
DR   Pfam; PF01002; Flavi_NS2B; 1.
DR   Pfam; PF01350; Flavi_NS4A; 1.
DR   Pfam; PF01349; Flavi_NS4B; 1.
DR   Pfam; PF00972; Flavi_NS5; 1.
DR   Pfam; PF01570; Flavi_propep; 1.
DR   Pfam; PF01728; FtsJ; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   ProDom; PD001556; Flavi_glycoprotE; 1.
DR   ProDom; PD001496; Flavi_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00690; DEAH_ATP_HELICASE; FALSE_NEG.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Hydrolase; Helicase;
KW   ATP-binding; Transmembrane; Nonstructural protein; 3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    112       CAPSID PROTEIN C.
FT   PROPEP      113    205
FT   CHAIN       206    280       ENVELOPE GLYCOPROTEIN M.
FT   CHAIN       281    776       MAJOR ENVELOPE PROTEIN E.
FT   CHAIN       777   1128       NONSTRUCTURAL PROTEIN NS1.
FT   CHAIN      1129   1358       NONSTRUCTURAL PROTEIN NS2A.
FT   CHAIN      1359   1489       NONSTRUCTURAL PROTEIN NS2B.
FT   CHAIN      1490   2110       PROTEASE/HELICASE (NS3).
FT   CHAIN      2111   2259       NONSTRUCTURAL PROTEIN NS4A.
FT   CHAIN      2260   2511       NONSTRUCTURAL PROTEIN NS4B.
FT   CHAIN      2512   3414       RNA-DIRECTED RNA POLYMERASE (NS5).
FT   DOMAIN      378    391       INVOLVED IN FUSION.
FT   NP_BIND    1688   1695       ATP (POTENTIAL).
FT   SITE       1779   1782       DEAH BOX.
FT   TRANSMEM    101    112       HYDROPHOBIC SIGNAL SEQUENCE (POTENTIAL).
FT   TRANSMEM    247    259       POTENTIAL.
FT   TRANSMEM    266    280       POTENTIAL.
FT   TRANSMEM    738    751       POTENTIAL.
FT   DISULFID    283    310
FT   DISULFID    340    396
FT   DISULFID    354    385
FT   DISULFID    372    401
FT   DISULFID    466    570
FT   DISULFID    587    618
FT   CARBOHYD    144    144       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    434    434       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    861    861       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    983    983       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    999    999       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2447   2447       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   HELIX       282    285
FT   STRAND      290    294
FT   STRAND      300    306
FT   TURN        307    308
FT   STRAND      311    315
FT   TURN        316    317
FT   STRAND      318    331
FT   STRAND      334    340
FT   STRAND      343    352
FT   TURN        355    356
FT   HELIX       363    366
FT   TURN        368    369
FT   STRAND      370    379
FT   HELIX       381    383
FT   TURN        384    384
FT   STRAND      389    401
FT   TURN        403    404
FT   STRAND      406    411
FT   TURN        414    416
FT   STRAND      418    425
FT   TURN        433    434
FT   TURN        438    439
FT   STRAND      441    446
FT   TURN        447    448
FT   STRAND      449    449
FT   STRAND      451    455
FT   HELIX       457    459
```

```
FT   STRAND      460    467
FT   HELIX       468    470
FT   TURN        475    476
FT   STRAND      477    481
FT   TURN        484    485
FT   TURN        487    488
FT   STRAND      492    496
FT   HELIX       497    501
FT   TURN        502    502
FT   STRAND      507    508
FT   TURN        510    511
FT   STRAND      516    516
FT   TURN        517    517
FT   HELIX       518    521
FT   STRAND      522    524
FT   STRAND      528    528
FT   TURN        529    530
FT   STRAND      531    531
FT   STRAND      534    536
FT   HELIX       541    547
FT   TURN        548    550
FT   STRAND      553    557
FT   TURN        558    559
FT   STRAND      560    562
FT   STRAND      567    573
FT   TURN        575    576
FT   STRAND      577    577
FT   TURN        582    583
FT   STRAND      586    586
```

Query Match 5.0%; Score 88; DB 1; Length 3414;
Best Local Similarity 23.0%; Pred. No. 59;
Matches 59; Conservative 28; Mismatches 86; Indels 84; Gaps 12;

```
Qy          64 HLTPLQD-------WAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSA 116
               |||  |:        |  |||     |   : :| :   :  |           | |  |
Db        1442 HLTELEKEERMMAFWLIAGL-----AASAIHWSGILGVMGLWTLTEMLRSSRRSDLVFSG 1496

Qy         117 RRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVST 176
               : |||    | :  :|  :|: :|   : |                |||       |
Db        1497 QGGRERGDRPPEVKDGV-YRIFSPGLFWGQ----------------NQVG-----VGY 1532

Qy         177 ATQSFLATCVNGVCWTVFHGAG---SKTLAGPKGPITQMYTNVDQDLV----------GW 223
                ::  | |       | |  ||       :|||       : :| :|:|          |
Db        1533 GSKGVLHT-----MWHVTRGAALSIDDAVAGP------YWADVREDVVCYGGAWSLEEKW 1581

Qy         224 QA--------PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKG 275
               :          ||| |:     |   :| | |               |  |   |:  :||
Db        1582 KGETVQVHAFPPG-RAHEVHQCQPGELILDT--------------GRKLGAIPIDLVKG 1625

Qy         276 SSGGPLLCPSGHAVGIF 292
               :|| |:|   |  ||::
Db        1626 TSGSPILNAQGVVVGLY 1642
```

Search completed: May 6, 2004, 09:31:48
Job time : 9.97286 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:21:36 ; Search time 31.2384 Seconds
(without alignments)
3373.509 Million cell updates/sec

Title: US-10-650-585-4
Perfect score: 1771
Sequence: 1 MKKKKLEHHHHHHTSAGITK.....TTMRTSSAWRHPQFGGKKKK 334

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1551	87.6	3010	12	Q9J3H7	Q9j3h7 hepatitis c
2	1546	87.3	3010	12	Q68826	Q68826 hepatitis c
3	1546	87.3	3010	12	P90191	P90191 hepatitis c
4	1545	87.2	3010	12	Q9DTE6	Q9dte6 hepatitis c
5	1543	87.1	3010	12	Q9DTE4	Q9dte4 hepatitis c
6	1542	87.1	3010	12	Q9DTD6	Q9dtd6 hepatitis c
7	1541	87.0	3010	12	P88803	P88803 hepatitis c
8	1540	87.0	3010	12	Q9J3H5	Q9j3h5 hepatitis c
9	1540	87.0	3010	12	Q807P3	Q807p3 hepatitis c
10	1538	86.8	3010	12	Q9J3F9	Q9j3f9 hepatitis c
11	1535	86.7	361	12	O70815	O70815 hepatitis c
12	1535	86.7	3008	12	Q9J3F4	Q9j3f4 hepatitis c
13	1535	86.7	3010	12	Q9J3H3	Q9j3h3 hepatitis c
14	1534	86.6	3010	12	Q9J3H2	Q9j3h2 hepatitis c
15	1534	86.6	3010	12	Q9J3I0	Q9j3i0 hepatitis c
16	1533	86.6	3010	12	Q9QIY3	Q9qiy3 hepatitis c
17	1532	86.5	3010	12	Q9QIY7	Q9qiy7 hepatitis c
18	1532	86.5	3010	12	Q9QIY8	Q9qiy8 hepatitis c
19	1532	86.5	3010	12	Q9QIX6	Q9qix6 hepatitis c
20	1532	86.5	3010	12	Q9QIX5	Q9qix5 hepatitis c
21	1532	86.5	3010	12	Q9J3H6	Q9j3h6 hepatitis c
22	1532	86.5	3010	12	O09796	O09796 hepatitis c
23	1531	86.4	3010	12	Q91AU0	Q91au0 hepatitis c
24	1531	86.4	3010	12	Q9J3H9	Q9j3h9 hepatitis c
25	1531	86.4	3010	12	Q9WMX2	Q9wmx2 hepatitis c
26	1530	86.4	3013	12	Q9J3H4	Q9j3h4 hepatitis c
27	1529	86.3	1186	12	Q81755	Q81755 hepatitis c
28	1529	86.3	2284	12	Q81817	Q81817 hepatitis c
29	1529	86.3	3010	12	Q68788	Q68788 hepatitis c
30	1529	86.3	3010	12	P89966	P89966 hepatitis c
31	1529	86.3	3010	12	Q9DTD7	Q9dtd7 hepatitis c
32	1529	86.3	3010	12	Q9DTF0	Q9dtf0 hepatitis c
33	1529	86.3	3014	12	Q9DTE0	Q9dte0 hepatitis c
34	1528	86.3	3010	12	Q99AU2	Q99au2 hepatitis c
35	1528	86.3	3011	12	Q9DTE3	Q9dte3 hepatitis c
36	1527	86.2	361	12	O70818	O70818 hepatitis c
37	1527	86.2	3010	12	Q9QP61	Q9qp61 hepatitis c
38	1527	86.2	3010	12	Q9J3G6	Q9j3g6 hepatitis c
39	1526	86.2	3010	12	Q8QRL8	Q8qrl8 hepatitis c
40	1525	86.1	3010	12	Q9J3I1	Q9j3i1 hepatitis c
41	1524	86.1	3010	12	Q9QIY5	Q9qiy5 hepatitis c
42	1524	86.1	3010	12	Q9DTE9	Q9dte9 hepatitis c
43	1524	86.1	3010	12	Q9QIY4	Q9qiy4 hepatitis c
44	1524	86.1	3010	12	Q81541	Q81541 hepatitis c
45	1523	86.0	3010	12	Q9J3G3	Q9j3g3 hepatitis c

ALIGNMENTS

RESULT 1
Q9J3H7
ID Q9J3H7 PRELIMINARY; PRT; 3010 AA.
AC Q9J3H7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD15;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
RT progression.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE; THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF207756; AAF65946.1; -.
DR PIR; A61196; A61196.
DR PIR; PQ0246; PQ0246.
DR PIR; PQ0804; PQ0804.
DR PIR; PS0329; PS0329.
DR HSSP; P26663; 1JXP.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.

```
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327365 MW;  D8653F7317FFA106 CRC64;

  Query Match             87.6%;  Score 1551;  DB 12;  Length 3010;
  Best Local Similarity   95.4%;  Pred. No. 4.6e-126;
  Matches  290;  Conservative   10;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||::|||||||||||||||||||||:|||||||||||||||:|||| ||| |||||
Db        904 AGITRMPYFVRAQGLIRACMLVRKAAGGHYIQMAFMKLAALTGTYIYDHLAPLQHWAHAG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||:|||| |||||||||||||| ||||||||||:||||||:||||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREVLLGPADSFEGQGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCINGVCWTVFH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        316 TMRT 319
              ||||:
Db       1204 TMRS 1207


RESULT 2
Q68825
ID   Q68826      PRELIMINARY;      PRT;  3010 AA.
AC   Q68826;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J33;
RA   Cho M.J.;
RT   "Molecular cloning of Hepatitis C virus genome from a single Japanese
RT   patient.";
RL   Submitted (SEP-1991) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; D14484; BAA03375.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PQ0804; PQ0804.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW   Hydrolase; Nonstructural protein; Polyprotein;
KW   RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327150 MW;  7270F47984554FAD CRC64;

  Query Match             87.3%;  Score 1546;  DB 12;  Length 3010;
  Best Local Similarity   96.1%;  Pred. No. 1.3e-125;
```

```
  Matches  292;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              |||| |||||||||||||||||| |||||||||||||||||||||||||:||||||
Db        904 AGITAVPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLRDWAHAG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||:||||| |||||||||||||| |||||||||||||||: |||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGPADSLEGQGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              |||||||||||||||||||||||||||||||||||||||||||||||||:|||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCINGVCWTVYH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 RRGDGRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203

Qy        316 TMRT 319
              |||:
Db       1204 TMRS 1207


RESULT 3
P90191
ID   P90191      PRELIMINARY;      PRT;  3010 AA.
AC   P90191;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCV-1b;
RA   Enomoto N.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCV-1b;
RX   MEDLINE=95340824; PubMed=7542279;
RA   Enomoto N., Sakuma I., Asahina Y., Kurosaki M., Murakami T.,
RA   Yamamoto C., Izumi N., Marumo F., Sato C.;
RT   "Comparison of full-length sequences of interferon-sensitive and
RT   resistant hepatitis C virus 1b.";
RL   J. Clin. Invest. 96:224-230(1995).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; D50482; BAA09073.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0254; PQ0254.
DR   PIR; PQ0804; PQ0804.
DR   PIR; PS0329; PS0329.
DR   PDB; 1DXW; 12-JAN-01.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
FT   CHAIN         1    191       CORE PROTEIN.
FT   CHAIN       192    383       E1.
FT   CHAIN       384    809       E2.
FT   CHAIN       810   1026       NS2.
FT   CHAIN      1027   1657       NS3.
FT   CHAIN      1658   1711       NS4A.
FT   CHAIN      1712   1972       NS4B.
FT   CHAIN      1973   2419       NS5A.
FT   CHAIN      2420   3010       NS5B.
SQ   SEQUENCE   3010 AA;  327438 MW;  5F15AC675A0C8268 CRC64;

  Query Match             87.3%;  Score 1546;  DB 12;  Length 3010;
  Best Local Similarity   95.7%;  Pred. No. 1.3e-125;
  Matches  291;  Conservative    9;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||:|||||||||||||||||| |||||||||:||||||||||:|||||||||| |
Db        904 AGLTKVPYFVRAQGLIRACMLVRKVAGGHYVQMAFVKLAALTGTYVYNHLTPLQDWAHTG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||:||||| |||||||||||||| |||||||||||||||:||||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIISGLAVSARRGREILLGPADSFEGQGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        316 TMRT 319
              |||:
Db       1204 TMRS 1207
```

```
RESULT 4
Q9DTE6
ID   Q9DTE6      PRELIMINARY;      PRT;  3010 AA.
AC   Q9DTE6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCVT142;
RA   Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA   Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA   Mishiro S.;
RT   "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT   with hepatocellular carcinoma: the 'progression score' revisited.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AB049091; BAB18804.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW   Hydrolase; Nonstructural protein; Polyprotein;
KW   RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327042 MW;  3807DC6879684C95 CRC64;

  Query Match           87.2%;  Score 1545;  DB 12;  Length 3010;
  Best Local Similarity  95.4%;  Pred. No. 1.6e-125;
  Matches  290;  Conservative    9;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||:|||||||||||||||||| |||||||||||||||||||||||:|||| |
Db        904 AGITRVPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLRDWAHTG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||:||||| |||||||||||||||||||||||||||||||: |||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIISGLPVSARRGREILLGPADSLEGQGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||| |||||||||||||:|||:|||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHVVGIFRAAVCTRGVAKAVDFVPVEAMET 1203

Qy        316 TMRT 319
              |||:
Db       1204 TMRS 1207


RESULT 5
Q9DTE4
ID   Q9DTE4      PRELIMINARY;      PRT;  3010 AA.
AC   Q9DTE4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCVT150;
RA   Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA   Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA   Mishiro S.;
RT   "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT   with hepatocellular carcinoma: the 'progression score' revisited.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AB049093; BAB18806.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PQ0804; PQ0804.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
```

```
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327324 MW;  3DE6CF249BD1151C CRC64;

  Query Match            87.1%;  Score 1543;  DB 12;  Length 3010;
  Best Local Similarity  95.4%;  Pred. No. 2.3e-125;
  Matches  290;  Conservative    8;  Mismatches    6;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        904 AGITKVPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAS 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              |||||||||||:||||| |||||||||||||||| |||||||||:||||||: |||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGQGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCINGVCWTVYH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              |||| |||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       1144 RRGDGRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        316 TMRT 319
              ||||
Db       1204 TMRS 1207


RESULT 6
Q9DTD6
ID   Q9DTD6      PRELIMINARY;      PRT;  3010 AA.
AC   Q9DTD6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCVT221;
RA   Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA   Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA   Mishiro S.;
RT   "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT   with hepatocellular carcinoma: the 'progression score' revisited.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AB049101; BAB18814.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
```

```
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW   Hydrolase; Nonstructural protein; Polyprotein;
KW   RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327108 MW;  DE182D810EF78EE4 CRC64;

  Query Match            87.1%;  Score 1542;  DB 12;  Length 3010;
  Best Local Similarity  95.7%;  Pred. No. 2.8e-125;
  Matches  291;  Conservative    7;  Mismatches    6;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              | | |||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        904 AVIIKVPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              |||||||||||:||||| ||||||||||||||||||||||||||||||:||||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIISGLPVSARRGREILLGPADSFEGQGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db       1024 RLLAPITAYSQQTRGLLGCIVTSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              |||||||||||||||||||||||||||| |||||||:||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWPAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||:||||||||||||||||||||||||||| ||:||||||||||||||||||||||||
Db       1144 RRGDTRGSLLSPRPVSYLKGSSGGPLLCPSGHIVGVFRAAVCTRGVAKAVDFIPVESMET 1203

Qy        316 TMRT 319
              |||:
Db       1204 TMRS 1207


RESULT 7
P88803
ID   P88803      PRELIMINARY;      PRT;  3010 AA.
AC   P88803;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCV-1b;
RA   Enomoto N.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCV-1b;
RX   MEDLINE=95340824; PubMed=7542279;
RA   Enomoto N., Sakuma I., Asahina Y., Kurosaki M., Murakami T.,
RA   Yamamoto C., Izumi N., Marumo F., Sato C.;
RT   "Comparison of full-length sequences of interferon-sensitive and
RT   resistant hepatitis C virus 1b.";
RL   J. Clin. Invest. 96:224-230(1995).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; D50484; BAA09075.1; -.
DR   PIR; A61196; A61196.
DR   HSSP; P26663; 1NS3.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
FT   CHAIN         1    191       CORE PROTEIN.
FT   CHAIN       192    383       E1.
FT   CHAIN       384    809       E2.
FT   CHAIN       810   1026       NS2.
FT   CHAIN      1027   1657       NS3.
FT   CHAIN      1658   1711       NS4A.
FT   CHAIN      1712   1972       NS4B.
FT   CHAIN      1973   2419       NS5A.
FT   CHAIN      2420   3010       NS5B.
SQ   SEQUENCE   3010 AA;  327332 MW;  5F81505783FEFFB8 CRC64;

  Query Match            87.0%;  Score 1541;  DB 12;  Length 3010;
  Best Local Similarity  95.1%;  Pred. No. 3.5e-125;
  Matches  289;  Conservative   10;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||: :|||||||||||| |||||| |||||||||||:||||||||||:|||||||||| |
Db        904 AGLIRVPYFVRAQGLIRVCMLVRKVAGGHYVQMAPVKLAALTGTYVYNHLTPLQDWAHTG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              |||||||||||:||||| ||||||||||||||||||||||||||||||:||||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIISGLPVSARRGREILLGPADSFEGQGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
```

```
                 ||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
                 ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        316 TMRT 319
                 |||:
Db       1204 TMRS 1207


RESULT 8
Q9J3H5
ID   Q9J3H5      PRELIMINARY;      PRT;  3010 AA.
AC   Q9J3H5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD17;
RA   Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT   "Characteristics of hepatitis C viral genome associated with disease
RT   progression.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF207758; AAF65948.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PQ0254; PQ0254.
DR   PIR; PS0329; PS0329.
DR   HSSP; P27958; 1HEI.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW   Hydrolase; Nonstructural protein; Polyprotein;
KW   RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  326801 MW;  9FEE3D1B93B7AA4B CRC64;

  Query Match           87.0%;  Score 1540;  DB 12;  Length 3010;
  Best Local Similarity  95.4%;  Pred. No. 4.2e-125;
  Matches  290;  Conservative    8;  Mismatches    6;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||:|||||||||||||||||||| |||||||| |||||||||||||||||||:|||| |
Db        904 AGITRVPYFVRAQGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHTG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||:||||| ||||||||||||||||||||||||||||||: |||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIISGLPVSARRGREILLGPADSLEGQGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCINGVCWTVYH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDLDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGLFRAAVCTRGVAKAVDFIPVESMET 1203

Qy        316 TMRT 319
              |||:
Db       1204 TMRS 1207


RESULT 9
Q807P3
ID   Q807P3      PRELIMINARY;      PRT;  3010 AA.
AC   Q807P3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M1LE;
RX   MEDLINE=22047193; PubMed=12051758;
RA   Kishine H., Sugiyama K., Hijikata M., Kato N., Takahashi H., Noshi T.,
RA   Nio Y., Hosaka M., Miyanari Y., Shimotohno K.;
RT   "Subgenomic replicon derived from a cell line infected with the
RT   hepatitis C virus.";
RL   Biochem. Biophys. Res. Commun. 293:993-999(2002).
DR   EMBL; AB080299; BAC54896.1; -.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
```

```
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Polyprotein.
SQ   SEQUENCE   3010 AA;  327097 MW;  EE6418C7A723E686 CRC64;

  Query Match           87.0%;  Score 1540;  DB 12;  Length 3010;
  Best Local Similarity  95.7%;  Pred. No. 4.2e-125;
  Matches  291;  Conservative    6;  Mismatches    7;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||:|||||||||||||||||||||||||||||||:||||||||||||| ||| ||| |
Db        904 AGITRVPYFVRAQGLIRACMLVRKAAGGHYVQMAFVKLAALTGTYVYDHLAPLQHWAHTG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||:|||| |||||||||||||| ||||||||||||||||: |||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGPADSLEGQGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCINGVCWTVFH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 RRGDGRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203

Qy        316 TMRT 319
              |||:
Db       1204 TMRS 1207
```

```
RESULT 10
Q9J3F9
ID   Q9J3F9      PRELIMINARY;      PRT;  3010 AA.
AC   Q9J3F9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD33;
RA   Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT   "Characteristics of hepatitis C viral genome associated with disease
RT   progression.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF207774; AAF65964.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PS0329; PS0329.
DR   HSSP; P27958; 1HEI.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
```

```
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327102 MW;  7162C9DB93E6E0C7 CRC64;

  Query Match             86.8%;  Score 1538;  DB 12;  Length 3010;
  Best Local Similarity   94.7%;  Pred. No. 6.4e-125;
  Matches  288;  Conservative   11;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||::|||||||||||||||||| |||||||||||| |||||:|||||:||||||
Db        904 AGITRMPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLGALTGTYIYDHLTPLRDWAHAG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              |||||||||||:|||| |||||||||||||||||||||:||||||||: ||:||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIISGLPVSARRGKEILLGPADSLEGRGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              |||||||||||||||||||||||||||| |||||||:|||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWPAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        316 TMRT 319
              ||||:
Db       1204 TMRS 1207


RESULT 11
O70815
ID   O70815      PRELIMINARY;      PRT;   361 AA.
AC   O70815;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Polyprotein (Fragment).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98321154; PubMed=9656998;
RA   Yamada K., Mori A., Seki M., Kimura J., Yuasa S., Matsuura Y.,
RA   Miyamura T.;
RT   "Critical point mutations for hepatitis C virus NS3 proteinase.";
RL   Virology 246:104-112(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Mori A., Yamada K., Kimura J., Koide T., Yuasa S., Yamada E.,
RA   Miyamura T.;
RT   "Enzymatic characterization of purified NS3 serine proteinase of
RT   hepatitis C virus expressed in Escherichia coli.";
RL   FEBS Lett. 378:37-42(1998).
DR   EMBL; AB013620; BAA28498.1; -.
DR   HSSP; P27958; 1HEI.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; Peptidase_C29.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
FT   NON_TER       1      1
FT   NON_TER     361    361
SQ   SEQUENCE   361 AA;  38336 MW;  87DC310C76F4BCC3 CRC64;

  Query Match             86.7%;  Score 1535;  DB 12;  Length 361;
  Best Local Similarity   94.7%;  Pred. No. 7.1e-126;
  Matches  288;  Conservative   10;  Mismatches    6;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||:|||||||||||||||||| |||||||||||||||||||||||||||||||||
Db          5 AGITRVPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHVS 64

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              |||||||||||:|||| |||||||||||||||||||||||||:|||||:|| |||
Db         65 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIISGLPVSARRGREVLLGPADSFEHQGW 124

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWSVYH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|:|
Db        125 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWSVYH 184

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              |||||||||||:|||||||| |||||||||||||:||||||||||||||||||||||||
Db        185 GAGSKTLAGPKGPVTQMYTNVDLDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 244

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db        245 RRGDSRGSLLSPRPISYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 304

Qy        316 TMRT 319
              ||||:
Db        305 TMRS 308


RESULT 12
Q9J3F4
ID   Q9J3F4      PRELIMINARY;      PRT;  3008 AA.
AC   Q9J3F4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
GN   MD34.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD34;
RA   Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT   "Characteristics of hepatitis C viral genome associated with disease
RT   progression.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF208024; AAF61205.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
```

```
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW   Hydrolase; Nonstructural protein; Polyprotein;
KW   RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3008 AA;  326834 MW;  99AE09E14C3109F4 CRC64;

  Query Match             86.7%;  Score 1535;  DB 12;  Length 3008;
  Best Local Similarity   95.4%;  Pred. No. 1.2e-124;
  Matches  290;  Conservative    6;  Mismatches    8;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              | | |||||||||||||||| || |||||||||||||||||||||||||||||||||||
Db        902 ASIIKVPYFVRAQGLIRACMLARKVAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAS 961

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              |||||||||||:||||| |||||||||||||| |||||||||:||||||||: |||||
Db        962 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGQGW 1021

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1022 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 1081

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              |||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db       1082 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1141

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       1142 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1201

Qy        316 TMRT 319
              ||||:
Db       1202 TMRS 1205


RESULT 13
Q9J3H3
```

```
ID   Q9J3H3      PRELIMINARY;      PRT;  3010 AA.
AC   Q9J3H3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD19;
RA   Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT   "Characteristics of hepatitis C viral genome associated with disease
RT   progression.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF207760; AAF65950.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
```

```
SQ   SEQUENCE   3010 AA;  327234 MW;  44C34677649CB8DD CRC64;

  Query Match             86.7%;  Score 1535;  DB 12;  Length 3010;
  Best Local Similarity   94.1%;  Pred. No. 1.2e-124;
  Matches  286;  Conservative   12;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
               ||||:|||||||||||||||:| || ||||||||||||||||||||||||||||||||||
Db         904 AGITRVPYFVRAQGLIRACLLARKVAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 963

Qy          76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
               ||||||||||:||||| |||||||||||||||| |||||||||:|| |||||| |||||
Db         964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEIFLGPADNLEGQGW 1023

Qy         136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
               ||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||:|
Db        1024 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy         196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
               |||:|||||||||||||||||||||:||||||:||||||||||||||||||||||||||
Db        1084 GAGAKTLAGPKGPITQMYTNVDQDLVGWQSPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy         256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
               ||||||||||||:|||||||||||||||||||||||||||||||||||||:|||||||||
Db        1144 RRGDSRGSLLSPRPISYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy         316 TMRT 319
               ||||:
Db        1204 TMRS 1207


RESULT 14
Q9J3H2
ID   Q9J3H2      PRELIMINARY;      PRT;  3010 AA.
AC   Q9J3H2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD20;
RA   Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT   "Characteristics of hepatitis C viral genome associated with disease
RT   progression.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF207761; AAF65951.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1NS3.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  326763 MW;  1A48EE4BE51440D0 CRC64;

  Query Match             86.6%;  Score 1534;  DB 12;  Length 3010;
  Best Local Similarity   94.7%;  Pred. No. 1.4e-124;
  Matches  288;  Conservative   10;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
               |||||||||||| |||||||||| |||||:|||||||||||||||||||:||||||
Db         904 AGITKVPYFVRAHGLIRACMLVRKVAGGHYLQMAFMKLAALTGTYVYDHLTPLRDWAHAG 963

Qy          76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
               ||||||||||:||||| |||||||||||||||||||||||||||||||||: |||||
Db         964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIISGLPVSARRGREILLLGPADSLEGQGW 1023

Qy         136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
               |||||||||:|||||||||||||||||||||||||||||||||||||||||||||||:|
Db        1024 RLLAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy         196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
               ||||| ||| |||:|||||||||||||||||||||:|||||||||||||||||||||||
Db        1084 GAGSKALAGQKGPVTQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy         256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
               ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy         316 TMRT 319
               ||||:
Db        1204 TMRS 1207


RESULT 15
Q9J3I0
ID   Q9J3I0      PRELIMINARY;      PRT;  3010 AA.
AC   Q9J3I0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
```

```
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD12;
RA   Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT   "Characteristics of hepatitis C viral genome associated with disease
RT   progression.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF207753; AAF65943.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW   Hydrolase; Nonstructural protein; Polyprotein;
KW   RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  326692 MW;  074098DB305AF1A9 CRC64;
```

Query Match 86.6%; Score 1534; DB 12; Length 3010;

Best Local Similarity 95.1%; Pred. No. 1.4e-124;
Matches 289; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

```
Qy        16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
             ||||:||||||||||||||||| |||||||||||||||||||||||||||||: ||| |
Db       904 AGITRVPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLRGWAHTG 963

Qy        76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
             ||||||||||:||||| |||||||||||||| |||||||||||||||||: ||||| 
Db       964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGPADSLEGQGW 1023

Qy       136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
             ||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||| 
Db      1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQTFLATCVNGVCWTVFH 1083

Qy       196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db      1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 1143

Qy       256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
             |||| |||||||||||||||||||||||| |||||||||||||||||:||||||| 
Db      1144 RRGDGRGSLLSPRPVSYLKGSSGGPLLCPSGHVVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy       316 TMRT 319
             |||: 
Db      1204 TMRS 1207
```

Search completed: May 6, 2004, 09:35:44
Job time : 34.2384 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:25:16 ; Search time 13.109 Seconds
(without alignments)
1315.364 Million cell updates/sec

Title: US-10-650-585-4
Perfect score: 1771
Sequence: 1 MKKKKLEHHHHHHTSAGITK.....TTMRTSSAWRHPQFGGKKKK 334

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1531	86.4	2201	4	US-09-539-601-6	Sequence 6, Appli
2	1531	86.4	2201	4	US-09-539-601-15	Sequence 15, Appl
3	1531	86.4	3010	4	US-09-539-601-3	Sequence 3, Appli
4	1531	86.4	3010	4	US-09-539-601-21	Sequence 21, Appl
5	1531	86.4	3010	4	US-09-539-601-27	Sequence 27, Appl
6	1528	86.3	1692	3	US-09-263-933-4	Sequence 4, Appli
7	1528	86.3	1692	4	US-09-919-901-4	Sequence 4, Appli
8	1528	86.3	2307	3	US-09-263-933-2	Sequence 2, Appli
9	1528	86.3	2307	4	US-09-919-901-2	Sequence 2, Appli
10	1525	86.1	1692	3	US-09-263-933-11	Sequence 11, Appl
11	1525	86.1	1692	4	US-09-919-901-11	Sequence 11, Appl
12	1525	86.1	2307	3	US-09-263-933-9	Sequence 9, Appli
13	1525	86.1	2307	4	US-09-919-901-9	Sequence 9, Appli
14	1524	86.1	3010	4	US-09-539-601-33	Sequence 33, Appl
15	1516	85.6	1692	3	US-09-263-933-18	Sequence 18, Appl
16	1516	85.6	1692	4	US-09-919-901-18	Sequence 18, Appl
17	1516	85.6	2307	3	US-09-263-933-16	Sequence 16, Appl
18	1516	85.6	2307	4	US-09-919-901-16	Sequence 16, Appl
19	1505	85.0	3010	3	US-09-014-416-3	Sequence 3, Appli
20	1479	83.5	2013	1	US-08-324-977-12	Sequence 12, Appl
21	1479	83.5	2013	2	US-08-384-616-12	Sequence 12, Appl
22	1479	83.5	2013	2	US-08-904-686A-12	Sequence 12, Appl
23	1479	83.5	2013	3	US-09-315-850-12	Sequence 12, Appl
24	1479	83.5	2201	4	US-08-952-981A-2	Sequence 2, Appli
25	1479	83.5	2620	1	US-08-324-977-32	Sequence 32, Appl
26	1479	83.5	2620	2	US-08-384-616-32	Sequence 32, Appl
27	1479	83.5	2620	2	US-08-904-686A-32	Sequence 32, Appl
28	1479	83.5	2620	3	US-09-315-850-32	Sequence 32, Appl
29	1479	83.5	2621	1	US-08-324-977-36	Sequence 36, Appl
30	1479	83.5	2621	2	US-08-384-616-36	Sequence 36, Appl
31	1479	83.5	2621	2	US-08-904-686A-36	Sequence 36, Appl
32	1479	83.5	2621	3	US-09-315-850-36	Sequence 36, Appl
33	1479	83.5	3010	1	US-08-324-977-2	Sequence 2, Appli
34	1479	83.5	3010	1	US-08-324-977-14	Sequence 14, Appl
35	1479	83.5	3010	2	US-08-384-616-2	Sequence 2, Appli
36	1479	83.5	3010	2	US-08-384-616-14	Sequence 14, Appl
37	1479	83.5	3010	2	US-08-904-686A-2	Sequence 2, Appli
38	1479	83.5	3010	2	US-08-904-686A-14	Sequence 14, Appl
39	1479	83.5	3010	3	US-09-315-850-2	Sequence 2, Appli
40	1479	83.5	3010	3	US-09-315-850-14	Sequence 14, Appl
41	1406	79.4	3012	3	US-08-811-566-2	Sequence 2, Appli
42	1406	79.4	3012	4	US-09-034-756-2	Sequence 2, Appli
43	1405	79.3	2894	2	US-08-466-975A-23	Sequence 23, Appl
44	1405	79.3	2894	2	US-08-391-671A-23	Sequence 23, Appl
45	1405	79.3	2894	3	US-08-467-902A-23	Sequence 23, Appl

ALIGNMENTS

RESULT 1
US-09-539-601-6
; Sequence 6, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartenschlager, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2201
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-539-601-6

Query Match 86.4%; Score 1531; DB 4; Length 2201;
Best Local Similarity 94.7%; Pred. No. 2e-143;
Matches 288; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

```
Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||||||||| |||||||||| ||||||||| ||||||||||||||||:||||||
Db         95 AGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAG 154

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||:||||| |:||||||||||||| |||||||||||| |||||: |||||
Db        155 LRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQGW 214

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||:|
Db        215 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 274

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db        275 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 334

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        335 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 394

Qy        316 TMRT 319
              ||||:
Db        395 TMRS 398
```

RESULT 2

```
US-09-539-601-15
; Sequence 15, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
;  APPLICANT: Bartenschlager, Ralf FW
;  TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
;  FILE REFERENCE: all sequences
;  CURRENT APPLICATION NUMBER: US/09/539,601C
;  CURRENT FILING DATE: 2001-08-30
;  EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
;  EARLIER FILING DATE: 1999-04-03
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
;   LENGTH: 2201
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
US-09-539-601-15

  Query Match            86.4%;  Score 1531;  DB 4;  Length 2201;
  Best Local Similarity  94.7%;  Pred. No. 2e-143;
  Matches  288;  Conservative    9;  Mismatches    7;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              |||||||||||| |||||||||| |||||||| ||||||||||||||||||:||||||
Db         95 AGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAG 154

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              |||||||||||:||||| |:||||||||||||| |||||||||||| |||||: |||||
Db        155 LRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQGW 214

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||:|
Db        215 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 274

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db        275 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 334

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              |||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        335 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 394

Qy        316 TMRT 319
              |||:
Db        395 TMRS 398


RESULT 3
US-09-539-601-3
; Sequence 3, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
;  APPLICANT: Bartenschlager, Ralf FW
;  TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
;  FILE REFERENCE: all sequences
;  CURRENT APPLICATION NUMBER: US/09/539,601C
;  CURRENT FILING DATE: 2001-08-30
;  EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
;  EARLIER FILING DATE: 1999-04-03
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 3010
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
US-09-539-601-3

  Query Match            86.4%;  Score 1531;  DB 4;  Length 3010;
  Best Local Similarity  94.7%;  Pred. No. 3.2e-143;
  Matches  288;  Conservative    9;  Mismatches    7;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              |||||||||||| |||||||||| |||||||| ||||||||||||||||||:||||||
Db        904 AGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              |||||||||||:||||| |:||||||||||||| |||||||||||| |||||: |||||
Db        964 LRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              |||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        316 TMRT 319
              |||:
Db       1204 TMRS 1207


RESULT 4
US-09-539-601-21
; Sequence 21, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
;  APPLICANT: Bartenschlager, Ralf FW
;  TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
;  FILE REFERENCE: all sequences
;  CURRENT APPLICATION NUMBER: US/09/539,601C
;  CURRENT FILING DATE: 2001-08-30
;  EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
;  EARLIER FILING DATE: 1999-04-03
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
;   LENGTH: 3010
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
US-09-539-601-21

  Query Match            86.4%;  Score 1531;  DB 4;  Length 3010;
  Best Local Similarity  94.7%;  Pred. No. 3.2e-143;
  Matches  288;  Conservative    9;  Mismatches    7;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              |||||||||||| |||||||||| |||||||| ||||||||||||||||||:||||||
Db        904 AGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              |||||||||||:||||| |:||||||||||||| |||||||||||| |||||: |||||
Db        964 LRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              |||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        316 TMRT 319
              |||:
Db       1204 TMRS 1207
```

RESULT 5
US-09-539-601-27
; Sequence 27, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartenschlager, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 3010
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-539-601-27

Query Match 86.4%; Score 1531; DB 4; Length 3010;
Best Local Similarity 94.7%; Pred. No. 3.2e-143;
Matches 288; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

```
Qy          16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
               |||||||||||| |||||||||| ||||||||| ||||||||||||||||||:||||||
Db         904 AGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAG 963

Qy          76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
               ||||||||||:||||| |:|||||||||||||| |||||||||||| |||||: |||||
Db         964 LRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQGW 1023

Qy         136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
               |||||||||||||||||||||||||||||:|||||||||||||||||||||||||||:|
Db        1024 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy         196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
               |||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db        1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy         256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
               |||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy         316 TMRT 319
               ||||:
Db        1204 TMRS 1207
```

RESULT 6
US-09-263-933-4
; Sequence 4, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-4

Query Match 86.3%; Score 1528; DB 3; Length 1692;
Best Local Similarity 94.1%; Pred. No. 2.7e-143;
Matches 286; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

```
Qy          16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
               ||||:|||||||||| ||||||| |||||||||||||| ||||||:|:|||||:||||||
Db         183 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWAHAG 242

Qy          76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
               ||||||||||:||||| |||||||||||||| |||||||||:|||||||: ||:||
Db         243 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGRGW 302

Qy         136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db         303 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 362

Qy         196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
               |||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db         363 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 422

Qy         256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
               |||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db         423 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 482

Qy         316 TMRT 319
               ||||:
Db         483 TMRS 486
```

RESULT 7
US-09-919-901-4
; Sequence 4, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: :
US-09-919-901-4

Query Match 86.3%; Score 1528; DB 4; Length 1692;
Best Local Similarity 94.1%; Pred. No. 2.7e-143;
Matches 286; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

```
Qy          16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
               ||||:|||||||||| ||||||| |||||||||||||| ||||||:|:|||||:||||||
Db         183 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWAHAG 242

Qy          76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
               ||||||||||:||||| |||||||||||||| |||||||||:|||||||: ||:||
Db         243 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGRGW 302

Qy         136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db         303 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 362

Qy         196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
               |||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
```

```
Db        363 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 422

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        423 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 482

Qy        316 TMRT 319
              |||:
Db        483 TMRS 486


RESULT 8
US-09-263-933-2
; Sequence 2, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT: Jackson, Roberta L.
;  APPLICANT: Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/263,933
;  CURRENT FILING DATE: 1999-03-08
;  EARLIER APPLICATION NUMBER: 09/129,611
;  EARLIER FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 2307
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
US-09-263-933-2

  Query Match             86.3%;  Score 1528;  DB 3;  Length 2307;
  Best Local Similarity   94.1%;  Pred. No. 4.3e-143;
  Matches  286;  Conservative   12;  Mismatches    6;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||:|||||||||||| ||||||| ||||||||||||| ||||||:|:||||||:||||||
Db        275 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWAHAG 334

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||:||||| |||||||||||||||| |||||||||:||||||||: ||:||
Db        335 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGRGW 394

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        395 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 454

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db        455 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 514

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        515 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 574

Qy        316 TMRT 319
              |||:
Db        575 TMRS 578


RESULT 9
US-09-919-901-2
; Sequence 2, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT:  Jackson, Roberta L.
;  APPLICANT:  Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION:  OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/919,901
;  CURRENT FILING DATE:  2001-08-02
;  PRIOR APPLICATION NUMBER: 09/263,933
;  PRIOR FILING DATE: 1999-02-08
;  PRIOR APPLICATION NUMBER: 09/129,611
;  PRIOR FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 2307
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: :
US-09-919-901-2

  Query Match             86.3%;  Score 1528;  DB 4;  Length 2307;
  Best Local Similarity   94.1%;  Pred. No. 4.3e-143;
  Matches  286;  Conservative   12;  Mismatches    6;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||:|||||||||||| ||||||| ||||||||||||| ||||||:|:||||||:||||||
Db        275 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWAHAG 334

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||:||||| |||||||||||||||| |||||||||:||||||||: ||:||
Db        335 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGRGW 394

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        395 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 454

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db        455 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 514

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        515 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 574

Qy        316 TMRT 319
              |||:
Db        575 TMRS 578


RESULT 10
US-09-263-933-11
; Sequence 11, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT: Jackson, Roberta L.
;  APPLICANT: Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/263,933
;  CURRENT FILING DATE: 1999-03-08
;  EARLIER APPLICATION NUMBER: 09/129,611
;  EARLIER FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 1692
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
US-09-263-933-11

  Query Match             86.1%;  Score 1525;  DB 3;  Length 1692;
  Best Local Similarity   93.8%;  Pred. No. 5.4e-143;
  Matches  285;  Conservative   13;  Mismatches    6;  Indels    0;  Gaps    0;
```

```
Qy          16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
               ||||:|||||||||||| ||||||| |||||||||||||| ||||||:|:|||||:||||||
Db         183 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWAHAG 242

Qy          76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
               ||||||||||:|||||| |||||||||||||| |||||||||:|||||||: ||:||
Db         243 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGRGW 302

Qy         136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db         303 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 362

Qy         196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
               ||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db         363 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 422

Qy         256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
               ||||||||||||||||||||:|||||||||||||||||||||||||||||:|||||||
Db         423 RRGDSRGSLLSPRPVSYLKGSAGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 482

Qy         316 TMRT 319
               ||||:
Db         483 TMRS 486


RESULT 11
US-09-919-901-11
; Sequence 11, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT:  Jackson, Roberta L.
;  APPLICANT:  Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION:  OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/919,901
;  CURRENT FILING DATE:  2001-08-02
;  PRIOR APPLICATION NUMBER: 09/263,933
;  PRIOR FILING DATE: 1999-02-08
;  PRIOR APPLICATION NUMBER: 09/129,611
;  PRIOR FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 1692
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: :
US-09-919-901-11

  Query Match           86.1%;  Score 1525;  DB 4;  Length 1692;
  Best Local Similarity   93.8%;  Pred. No. 5.4e-143;
  Matches  285;  Conservative   13;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
               ||||:|||||||||||| ||||||| |||||||||||||| ||||||:|:|||||:||||||
Db         183 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWAHAG 242

Qy          76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
               ||||||||||:|||||| |||||||||||||| |||||||||:|||||||: ||:||
Db         243 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGRGW 302

Qy         136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db         303 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 362

Qy         196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
               ||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db         363 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 422

Qy         256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
               ||||||||||||||||||||:|||||||||||||||||||||||||||||:|||||||
Db         423 RRGDSRGSLLSPRPVSYLKGSAGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 482

Qy         316 TMRT 319
               ||||:
Db         483 TMRS 486


RESULT 12
US-09-263-933-9
; Sequence 9, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT: Jackson, Roberta L.
;  APPLICANT: Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/263,933
;  CURRENT FILING DATE: 1999-03-08
;  EARLIER APPLICATION NUMBER: 09/129,611
;  EARLIER FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 2307
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
US-09-263-933-9

  Query Match           86.1%;  Score 1525;  DB 3;  Length 2307;
  Best Local Similarity   93.8%;  Pred. No. 8.5e-143;
  Matches  285;  Conservative   13;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
               ||||:|||||||||||| ||||||| |||||||||||||| ||||||:|:|||||:||||||
Db         275 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWAHAG 334

Qy          76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
               ||||||||||:|||||| |||||||||||||| |||||||||:|||||||: ||:||
Db         335 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGRGW 394

Qy         136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db         395 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 454

Qy         196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
               ||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db         455 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 514

Qy         256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
               ||||||||||||||||||||:|||||||||||||||||||||||||||||:|||||||
Db         515 RRGDSRGSLLSPRPVSYLKGSAGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 574

Qy         316 TMRT 319
               ||||:
Db         575 TMRS 578


RESULT 13
US-09-919-901-9
; Sequence 9, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT:  Jackson, Roberta L.
;  APPLICANT:  Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION:  OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/919,901
;  CURRENT FILING DATE:  2001-08-02
```

```
;  PRIOR APPLICATION NUMBER: 09/263,933
;  PRIOR FILING DATE: 1999-02-08
;  PRIOR APPLICATION NUMBER: 09/129,611
;  PRIOR FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 2307
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: :
US-09-919-901-9

  Query Match             86.1%;  Score 1525;  DB 4;  Length 2307;
  Best Local Similarity   93.8%;  Pred. No. 8.5e-143;
  Matches  285;  Conservative   13;  Mismatches    6;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||:|||||||||||| ||||||| ||||||||||||| ||||||:|:||||||:||||||
Db        275 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWAHAG 334

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              |||||||||||:||||| |||||||||||||||| |||||||||:||||||||: ||:||
Db        335 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGRGW 394

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        395 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 454

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db        455 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 514

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              |||||||||||||||||||:|||||||||||||||||||||||||||||||:|||||||
Db        515 RRGDSRGSLLSPRPVSYLKGSAGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 574

Qy        316 TMRT 319
              |||:
Db        575 TMRS 578


RESULT 14
US-09-539-601-33
; Sequence 33, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
;  APPLICANT: Bartenschlager, Ralf FW
;  TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
;  FILE REFERENCE: all sequences
;  CURRENT APPLICATION NUMBER: US/09/539,601C
;  CURRENT FILING DATE: 2001-08-30
;  EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
;  EARLIER FILING DATE: 1999-04-03
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
;   LENGTH: 3010
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
US-09-539-601-33

  Query Match             86.1%;  Score 1524;  DB 4;  Length 3010;
  Best Local Similarity   94.4%;  Pred. No. 1.6e-142;
  Matches  287;  Conservative    9;  Mismatches    8;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              |||||||||||| |||||||||| ||||||||||||||||||:||||||
Db        904 AGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAG 963

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              |||||||||||:||||| |:|||||||||||||| |||||||||||: ||||||
```

```
Db        964 LRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQGW 1023

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:||||| |
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMGT 1203

Qy        316 TMRT 319
              |||:
Db       1204 TMRS 1207


RESULT 15
US-09-263-933-18
; Sequence 18, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT: Jackson, Roberta L.
;  APPLICANT: Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/263,933
;  CURRENT FILING DATE: 1999-03-08
;  EARLIER APPLICATION NUMBER: 09/129,611
;  EARLIER FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
;   LENGTH: 1692
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
US-09-263-933-18

  Query Match             85.6%;  Score 1516;  DB 3;  Length 1692;
  Best Local Similarity   93.4%;  Pred. No. 4.3e-142;
  Matches  284;  Conservative   13;  Mismatches    7;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||:|||||||||||| ||||||| ||||||||||||| ||||||:|:||||||:||||||
Db        183 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWAHAG 242

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              |||||||||||:||||| |||||||||||||||  |||||||||:||||||||: ||:||
Db        243 LRDLAVAVEPVVFSDMETKIITWGADTAAAGDIILGLPVSARRGKEILLGPADSLEGRGW 302

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        303 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 362

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db        363 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 422

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              |||||||||||||||||||:|||||||||||||||||||||||||||||||:|||||||
Db        423 RRGDSRGSLLSPRPVSYLKGSAGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 482

Qy        316 TMRT 319
              |||:
Db        483 TMRS 486


Search completed: May  6, 2004, 09:39:00
Job time : 14.109 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:30:56 ; Search time 34.167 Seconds
(without alignments)
2713.357 Million cell updates/sec

Title: US-10-650-585-4
Perfect score: 1771
Sequence: 1 MKKKKLEHHHHHHTSAGITK.....TTMRTSSAWRHPQFGGKKKK 334

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1771	100.0	334	13	US-10-017-736-4	Sequence 4, Appli
2	1771	100.0	334	16	US-10-650-585-4	Sequence 4, Appli
3	1660	93.7	409	13	US-10-017-736-2	Sequence 2, Appli
4	1660	93.7	409	16	US-10-650-585-2	Sequence 2, Appli
5	1589	89.7	303	13	US-10-017-736-10	Sequence 10, Appl
6	1589	89.7	303	16	US-10-650-585-10	Sequence 10, Appl
7	1589	89.7	341	13	US-10-017-736-14	Sequence 14, Appl
8	1589	89.7	341	16	US-10-650-585-14	Sequence 14, Appl
9	1589	89.7	352	13	US-10-017-736-13	Sequence 13, Appl
10	1589	89.7	352	16	US-10-650-585-13	Sequence 13, Appl
11	1589	89.7	380	13	US-10-017-736-12	Sequence 12, Appl
12	1589	89.7	380	16	US-10-650-585-12	Sequence 12, Appl
13	1589	89.7	393	13	US-10-017-736-11	Sequence 11, Appl
14	1589	89.7	393	16	US-10-650-585-11	Sequence 11, Appl
15	1580	89.2	303	13	US-10-017-736-18	Sequence 18, Appl
16	1580	89.2	303	16	US-10-650-585-18	Sequence 18, Appl
17	1579	89.2	303	13	US-10-017-736-16	Sequence 16, Appl
18	1579	89.2	303	16	US-10-650-585-16	Sequence 16, Appl
19	1570	88.7	301	13	US-10-017-736-17	Sequence 17, Appl
20	1570	88.7	301	16	US-10-650-585-17	Sequence 17, Appl
21	1532	86.5	292	13	US-10-017-736-15	Sequence 15, Appl
22	1532	86.5	292	16	US-10-650-585-15	Sequence 15, Appl
23	1531	86.4	2201	13	US-10-029-907-3	Sequence 3, Appli
24	1531	86.4	2201	14	US-10-309-561-3	Sequence 3, Appli
25	1531	86.4	3010	12	US-10-467-000-1	Sequence 1, Appli
26	1528	86.3	1692	10	US-09-919-901-4	Sequence 4, Appli
27	1528	86.3	1692	14	US-10-191-966-4	Sequence 4, Appli
28	1528	86.3	2307	10	US-09-919-901-2	Sequence 2, Appli
29	1528	86.3	2307	14	US-10-191-966-2	Sequence 2, Appli
30	1525	86.1	1692	10	US-09-919-901-11	Sequence 11, Appl
31	1525	86.1	1692	14	US-10-191-966-11	Sequence 11, Appl
32	1525	86.1	2307	10	US-09-919-901-9	Sequence 9, Appli
33	1525	86.1	2307	14	US-10-191-966-9	Sequence 9, Appli
34	1516	85.6	1692	10	US-09-919-901-18	Sequence 18, Appl
35	1516	85.6	1692	14	US-10-191-966-18	Sequence 18, Appl
36	1516	85.6	2307	10	US-09-919-901-16	Sequence 16, Appl
37	1516	85.6	2307	14	US-10-191-966-16	Sequence 16, Appl
38	1479	83.5	2201	13	US-10-085-476-2	Sequence 2, Appli
39	1406	79.4	3011	9	US-09-742-659-4	Sequence 4, Appli
40	1406	79.4	3011	10	US-09-891-894-3	Sequence 3, Appli
41	1406	79.4	3011	14	US-10-184-150-3	Sequence 3, Appli
42	1406	79.4	3011	15	US-10-328-997-3	Sequence 3, Appli
43	1406	79.4	3012	9	US-09-238-076-2	Sequence 2, Appli
44	1406	79.4	3012	10	US-09-995-937-2	Sequence 2, Appli
45	1406	79.4	3012	10	US-09-917-563-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-10-017-736-4
; Sequence 4, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 334
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-4

Query Match 100.0%; Score 1771; DB 13; Length 334;
Best Local Similarity 100.0%; Pred. No. 8.2e-163;
Matches 334; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MKKKKLEHHHHHHTSAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKKKKLEHHHHHHTSAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTY 60

Qy         61 VYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGR 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGR 120

Qy        121 EILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQS 180

Qy        181 FLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 FLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSS 240

Qy        241 DLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 DLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRG 300

Qy        301 VAKAVDFIPVESMETTMRTSSAWRHPQFGGKKKK 334
              ||||||||||||||||||||||||||||||||||
Db        301 VAKAVDFIPVESMETTMRTSSAWRHPQFGGKKKK 334


RESULT 2
US-10-650-585-4
; Sequence 4, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/650,585
;  CURRENT FILING DATE:  2003-08-28
;  PRIOR APPLICATION NUMBER: US/10/017,736A
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 334
;   TYPE: PRT
;   ORGANISM: HCV
US-10-650-585-4

  Query Match             100.0%;  Score 1771;  DB 16;  Length 334;
  Best Local Similarity   100.0%;  Pred. No. 8.2e-163;
  Matches  334;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MKKKKLEHHHHHHTSAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTY 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKKKKLEHHHHHHTSAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTY 60

Qy         61 VYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGR 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGR 120

Qy        121 EILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQS 180

Qy        181 FLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSS 240

Qy        241 DLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 DLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRG 300

Qy        301 VAKAVDFIPVESMETTMRTSSAWRHPQFGGKKKK 334
              ||||||||||||||||||||||||||||||||||
Db        301 VAKAVDFIPVESMETTMRTSSAWRHPQFGGKKKK 334


RESULT 3
US-10-017-736-2
; Sequence 2, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/017,736
;  CURRENT FILING DATE:  2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 409
;   TYPE: PRT
;   ORGANISM: HCV
US-10-017-736-2

  Query Match             93.7%;  Score 1660;  DB 13;  Length 409;
  Best Local Similarity   100.0%;  Pred. No. 6e-152;
  Matches  315;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         95 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 154

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        155 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 214

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        215 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 274

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        275 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 334

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        335 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 394

Qy        316 TMRTSSAWRHPQFGG 330
              |||||||||||||||
Db        395 TMRTSSAWRHPQFGG 409


RESULT 4
US-10-650-585-2
; Sequence 2, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/650,585
;  CURRENT FILING DATE:  2003-08-28
;  PRIOR APPLICATION NUMBER: US/10/017,736A
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 409
;   TYPE: PRT
;   ORGANISM: HCV
US-10-650-585-2

  Query Match             93.7%;  Score 1660;  DB 16;  Length 409;
  Best Local Similarity   100.0%;  Pred. No. 6e-152;
  Matches  315;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         95 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 154

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        155 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 214

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        215 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 274

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        275 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 334

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        335 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 394

Qy        316 TMRTSSAWRHPQFGG 330
                |||||||||||||||
Db        395 TMRTSSAWRHPQFGG 409


RESULT 5
US-10-017-736-10
; Sequence 10, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/017,736
;  CURRENT FILING DATE:  2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
;   LENGTH: 303
;   TYPE: PRT
;   ORGANISM: HCV
US-10-017-736-10

  Query Match             89.7%;  Score 1589;  DB 13;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 3e-145;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300

Qy        316 TMR 318
                |||
Db        301 TMR 303


RESULT 6
US-10-650-585-10
; Sequence 10, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/650,585
;  CURRENT FILING DATE:  2003-08-28
;  PRIOR APPLICATION NUMBER: US/10/017,736A
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
;   LENGTH: 303
;   TYPE: PRT
;   ORGANISM: HCV
US-10-650-585-10

  Query Match             89.7%;  Score 1589;  DB 16;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 3e-145;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300

Qy        316 TMR 318
                |||
Db        301 TMR 303


RESULT 7
US-10-017-736-14
; Sequence 14, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/017,736
;  CURRENT FILING DATE:  2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
;   LENGTH: 341
;   TYPE: PRT
;   ORGANISM: HCV
US-10-017-736-14

  Query Match             89.7%;  Score 1589;  DB 13;  Length 341;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-145;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         39 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 98

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db         99 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 158

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 218

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 278

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 338

Qy        316 TMR 318
              |||
Db        339 TMR 341


RESULT 8
US-10-650-585-14
; Sequence 14, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/650,585
;  CURRENT FILING DATE:  2003-08-28
;  PRIOR APPLICATION NUMBER: US/10/017,736A
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
;   LENGTH: 341
;   TYPE: PRT
;   ORGANISM: HCV
US-10-650-585-14

  Query Match             89.7%;  Score 1589;  DB 16;  Length 341;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-145;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         39 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 98

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 158

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 218

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 278

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 338

Qy        316 TMR 318
              |||
Db        339 TMR 341


RESULT 9
US-10-017-736-13
; Sequence 13, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/017,736
;  CURRENT FILING DATE:  2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 352
;   TYPE: PRT
;   ORGANISM: HCV
US-10-017-736-13

  Query Match             89.7%;  Score 1589;  DB 13;  Length 352;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-145;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         50 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 109

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        110 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 169

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        170 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 229

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        230 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 289

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        290 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 349

Qy        316 TMR 318
              |||
Db        350 TMR 352


RESULT 10
US-10-650-585-13
; Sequence 13, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/650,585
;  CURRENT FILING DATE:  2003-08-28
;  PRIOR APPLICATION NUMBER: US/10/017,736A
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 352
;   TYPE: PRT
;   ORGANISM: HCV
US-10-650-585-13

  Query Match             89.7%;  Score 1589;  DB 16;  Length 352;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-145;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          50 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 109

Qy          76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         110 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 169

Qy         136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         170 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 229

Qy         196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         230 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 289

Qy         256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         290 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 349

Qy         316 TMR 318
                |||
Db         350 TMR 352


RESULT 11
US-10-017-736-12
; Sequence 12, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/017,736
;  CURRENT FILING DATE:  2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
;   LENGTH: 380
;   TYPE: PRT
;   ORGANISM: HCV
US-10-017-736-12

  Query Match             89.7%;  Score 1589;  DB 13;  Length 380;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-145;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          78 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 137

Qy          76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         138 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 197

Qy         136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         198 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 257

Qy         196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         258 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 317

Qy         256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         318 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 377

Qy         316 TMR 318
                |||
Db         378 TMR 380


RESULT 12
US-10-650-585-12
; Sequence 12, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/650,585
;  CURRENT FILING DATE:  2003-08-28
;  PRIOR APPLICATION NUMBER: US/10/017,736A
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
;   LENGTH: 380
;   TYPE: PRT
;   ORGANISM: HCV
US-10-650-585-12

  Query Match             89.7%;  Score 1589;  DB 16;  Length 380;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-145;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          78 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 137

Qy          76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         138 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 197

Qy         136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         198 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 257

Qy         196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         258 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 317

Qy         256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         318 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 377

Qy         316 TMR 318
                |||
Db         378 TMR 380


RESULT 13
US-10-017-736-11
; Sequence 11, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/017,736
;  CURRENT FILING DATE:  2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 393
;   TYPE: PRT
;   ORGANISM: HCV
US-10-017-736-11

  Query Match             89.7%;  Score 1589;  DB 13;  Length 393;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-145;
```

Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
Db         91 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 150

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
Db        151 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 210

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
Db        211 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 270

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
Db        271 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 330

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
Db        331 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 390

Qy        316 TMR 318
Db        391 TMR 393
```

RESULT 14
US-10-650-585-11
; Sequence 11, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/650,585
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-11

Query Match 89.7%; Score 1589; DB 16; Length 393;
Best Local Similarity 100.0%; Pred. No. 4.3e-145;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
Db         91 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 150

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
Db        151 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 210

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
Db        211 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 270

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
Db        271 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 330

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
Db        331 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 390

Qy        316 TMR 318
Db        391 TMR 393
```

RESULT 15
US-10-017-736-18
; Sequence 18, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 303
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-18

Query Match 89.2%; Score 1580; DB 13; Length 303;
Best Local Similarity 99.7%; Pred. No. 2.2e-144;
Matches 302; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
Db          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60

Qy         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135
Db         61 LRDLAVAVEPVIFSDMEVKIITWGADTAAAGDIISGLPVSARRGREILLGPADNFEGQGW 120

Qy        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195
Db        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180

Qy        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255
Db        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240

Qy        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
Db        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300

Qy        316 TMR 318
Db        301 TMR 303
```

Search completed: May 6, 2004, 09:43:18
Job time : 36.167 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:08:45 ; Search time 41.2434 Seconds
(without alignments)
2075.771 Million cell updates/sec

Title: US-10-650-585-10
Perfect score: 1589
Sequence: 1 AGITKVPYFVRAQGLIRACM.....RGVAKAVDFIPVESMETTMR 303

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1589	100.0	303	5	ABG32183	Abg32183 HCV prote
2	1589	100.0	334	5	ABG32182	Abg32182 HCV prote
3	1589	100.0	341	5	ABG32187	Abg32187 HCV prote
4	1589	100.0	352	5	ABG32186	Abg32186 HCV prote
5	1589	100.0	380	5	ABG32185	Abg32185 HCV prote
6	1589	100.0	393	5	ABG32184	Abg32184 HCV prote
7	1589	100.0	409	5	ABG32181	Abg32181 HCV prote
8	1580	99.4	303	5	ABG32191	Abg32191 HCV prote
9	1579	99.4	303	5	ABG32189	Abg32189 HCV prote
10	1570	98.8	301	5	ABG32190	Abg32190 HCV prote
11	1532	96.4	292	5	ABG32188	Abg32188 HCV prote
12	1530	96.3	2201	5	ABG30601	Abg30601 Hepatitis
13	1530	96.3	2201	5	ABG30591	Abg30591 Hepatitis
14	1530	96.3	2201	5	ABG30600	Abg30600 Hepatitis
15	1530	96.3	2201	5	ABG30581	Abg30581 Hepatitis
16	1530	96.3	2201	5	ABG30593	Abg30593 Hepatitis
17	1530	96.3	2201	5	ABG30582	Abg30582 Hepatitis
18	1530	96.3	2201	5	ABG30580	Abg30580 Hepatitis
19	1530	96.3	2201	5	ABG30587	Abg30587 Hepatitis
20	1530	96.3	2201	5	ABG30599	Abg30599 Hepatitis
21	1530	96.3	2201	5	ABG30594	Abg30594 Hepatitis
22	1530	96.3	2201	5	ABG30598	Abg30598 Hepatitis
23	1530	96.3	2201	5	ABG30595	Abg30595 Hepatitis
24	1530	96.3	3010	5	ABG32458	Abg32458 Hepatitis
25	1530	96.3	3010	5	ABG32459	Abg32459 Hepatitis
26	1530	96.3	3010	5	ABG32451	Abg32451 Hepatitis
27	1530	96.3	3010	5	ABG32455	Abg32455 Hepatitis
28	1530	96.3	3010	5	ABG32457	Abg32457 Hepatitis
29	1530	96.3	3010	5	ABG32460	Abg32460 Hepatitis
30	1530	96.3	3010	5	ABG32461	Abg32461 Hepatitis
31	1530	96.3	3010	5	ABG32454	Abg32454 Hepatitis
32	1530	96.3	3010	5	AAE20477	Aae20477 HCV-S1 fu
33	1530	96.3	3011	5	ABG32456	Abg32456 Hepatitis
34	1527	96.1	2201	5	ABG30586	Abg30586 Hepatitis
35	1527	96.1	2201	5	ABG30589	Abg30589 Hepatitis
36	1527	96.1	2201	5	ABG30583	Abg30583 Hepatitis
37	1527	96.1	2201	5	ABG30588	Abg30588 Hepatitis
38	1527	96.1	2307	3	AAY70064	Aay70064 Recombina
39	1527	96.1	3010	2	AAR68622	Aar68622 HCV prote
40	1527	96.1	3010	2	AAR82694	Aar82694 Partial H
41	1526	96.0	2201	5	ABG30590	Abg30590 Hepatitis
42	1524	95.9	2307	3	AAY70065	Aay70065 Recombina
43	1524	95.9	3010	5	ABG32452	Abg32452 Hepatitis
44	1523	95.8	2201	5	ABG30584	Abg30584 Hepatitis
45	1523	95.8	2201	5	ABG30602	Abg30602 Hepatitis

ALIGNMENTS

RESULT 1
ABG32183
ID ABG32183 standard; protein; 303 AA.
XX
AC ABG32183;
XX
DT 05-NOV-2002 (first entry)
XX
DE HCV protease NS2/3 truncation mutant 904-1206.
XX
KW HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;
KW chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW chaotropic agent; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO200248375-A2.
XX
PD 20-JUN-2002.
XX
PF 13-DEC-2001; 2001WO-CA001796.
XX
PR 15-DEC-2000; 2000US-0256031P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
XX
DR WPI; 2002-599511/64.
XX
PT Novel polypeptide for screening inhibitors of non-structural proteases
PT useful as therapeutic agents against hepatitis C virus, comprises full
PT length non-structural protease, or its truncation.
XX
PS Claim 39; Page 58-59; 67pp; English.
XX
CC The invention relates to an isolated polypeptide consisting of a full-
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC residue amino acid 810 to 906, or having a minimal amino acid sequence
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC NS2/3 protease. Also included are (1) a composition (C) comprising an
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC its truncation or a mutated sequence, where the protease is in a solution
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide

CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC protease, involving isolating the protease in the presence of a
CC chaotropic agent, refolding the isolated protease by contacting it with a
CC reducing agent, and LDAO in the presence of reduced concentration of the
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC protease, involving diluting refolded inactive NS2/3 protease in a medium
CC containing an activation detergent to induce auto-cleavage of the NS2/3
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC protease, involving incubating the active NS2/3 protease produced by M2
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC cleavage products or their fragments, and measuring the presence or
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC active NS2/3 protease, involving carrying out M3 in the presence of, or
CC absence of the potential inhibitor, comparing the amount of uncleaved
CC NS2/3 protease, cleavage products or their fragments. The protease is
CC useful for detailed biochemical characterisation of the enzymes and in
CC the development of in vitro assays for screening novel inhibitors of
CC NS2/3 protease which are useful as therapeutic agents against HCV
CC infection (which causes chronic liver disease, cirrhosis and end-stage
CC liver disease. M1 is useful for high level production of protease. The
CC present sequence represents the NS2/3 truncation mutant 904-1206
CC (numbered relative to the full length NS2/3 protein)
XX
SQ Sequence 303 AA;

Query Match 100.0%; Score 1589; DB 5; Length 303;
Best Local Similarity 100.0%; Pred. No. 5.1e-146;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300

Qy        301 TMR 303
              |||
Db        301 TMR 303
```

RESULT 2
ABG32182
ID ABG32182 standard; protein; 334 AA.
XX
AC ABG32182;
XX
DT 05-NOV-2002 (first entry)
XX
DE HCV protease NS2/3 truncation 4K-6H (904-1206)st-4K.
XX
KW HCV; enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;
KW chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW chaotropic agent; 4K-6H (904-1206)st-4K; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX

FH Key Location/Qualifiers
FT Peptide 1. .15
FT /note= "4-Lys/His tag"
FT Protein 16. .302
FT /note= "Truncated NS2/3 protease"
FT Peptide 319. .334
FT /note= "Streptavidin/4-Lys tag"
XX
PN WO200248375-A2.
XX
PD 20-JUN-2002.
XX
PF 13-DEC-2001; 2001WO-CA001796.
XX
PR 15-DEC-2000; 2000US-0256031P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
XX
DR WPI; 2002-599511/64.
DR N-PSDB; ABK90407.
XX
PT Novel polypeptide for screening inhibitors of non-structural proteases
PT useful as therapeutic agents against hepatitis C virus, comprises full
PT length non-structural protease, or its truncation.
XX
PS Claim 39; Fig 9B; 67pp; English.
XX
CC The invention relates to an isolated polypeptide consisting of a full-
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC residue amino acid 810 to 906, or having a minimal amino acid sequence
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC NS2/3 protease. Also included are (1) a composition (C) comprising an
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC its truncation or a mutated sequence, where the protease is in a solution
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC protease, involving isolating the protease in the presence of a
CC chaotropic agent, refolding the isolated protease by contacting it with a
CC reducing agent, and LDAO in the presence of reduced concentration of the
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC protease, involving diluting refolded inactive NS2/3 protease in a medium
CC containing an activation detergent to induce auto-cleavage of the NS2/3
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC protease, involving incubating the active NS2/3 protease produced by M2
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC cleavage products or their fragments, and measuring the presence or
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC active NS2/3 protease, involving carrying out M3 in the presence of, or
CC absence of the potential inhibitor, comparing the amount of uncleaved
CC NS2/3 protease, cleavage products or their fragments. The protease is
CC useful for detailed biochemical characterisation of the enzymes and in
CC the development of in vitro assays for screening novel inhibitors of
CC NS2/3 protease which are useful as therapeutic agents against HCV
CC infection (which causes chronic liver disease, cirrhosis and end-stage
CC liver disease. M1 is useful for high level production of protease. The
CC present sequence represents the NS2/3 truncation mutant 4K-6H (904-
CC 1206)st-4K comprising a truncated NS2/3 protein with a four Lys/six His N
CC -terminal tag, a C-terminal streptavidin tag and C-terminal four Lys tag
XX
SQ Sequence 334 AA;

Query Match 100.0%; Score 1589; DB 5; Length 334;
Best Local Similarity 100.0%; Pred. No. 5.8e-146;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75
```

```
Qy          61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135

Qy         121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195

Qy         181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255

Qy         241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315

Qy         301 TMR 303
               |||
Db         316 TMR 318
```

RESULT 3
ABG32187
ID ABG32187 standard; protein; 341 AA.
XX
AC ABG32187;
XX
DT 05-NOV-2002 (first entry)
XX
DE HCV protease NS2/3 truncation mutant 866-1206.
XX
KW HCV; enzyme; protease; NS2/3 (866-1206); hepatitis C virus infection;
KW chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW chaotropic agent; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO200248375-A2.
XX
PD 20-JUN-2002.
XX
PF 13-DEC-2001; 2001WO-CA001796.
XX
PR 15-DEC-2000; 2000US-0256031P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
XX
DR WPI; 2002-599511/64.
XX
PT Novel polypeptide for screening inhibitors of non-structural proteases
PT useful as therapeutic agents against hepatitis C virus, comprises full
PT length non-structural protease, or its truncation.
XX
PS Claim 41; Page 62-63; 67pp; English.
XX
CC The invention relates to an isolated polypeptide consisting of a full-
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC residue amino acid 810 to 906, or having a minimal amino acid sequence
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC NS2/3 protease. Also included are (1) a composition (C) comprising an
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC its truncation or a mutated sequence, where the protease is in a solution
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC protease, involving isolating the protease in the presence of a
CC chaotropic agent, refolding the isolated protease by contacting it with a
CC reducing agent, and LDAO in the presence of reduced concentration of the
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC protease, involving diluting refolded inactive NS2/3 protease in a medium
CC containing an activation detergent to induce auto-cleavage of the NS2/3
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC protease, involving incubating the active NS2/3 protease produced by M2
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC cleavage products or their fragments, and measuring the presence or
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC active NS2/3 protease, involving carrying out M3 in the presence of, or
CC absence of the potential inhibitor, comparing the amount of uncleaved
CC NS2/3 protease, cleavage products or their fragments. The protease is
CC useful for detailed biochemical characterisation of the enzymes and in
CC the development of in vitro assays for screening novel inhibitors of
CC NS2/3 protease which are useful as therapeutic agents against HCV
CC infection (which causes chronic liver disease, cirrhosis and end-stage
CC liver disease. M1 is useful for high level production of protease. The
CC present sequence represents the NS2/3 truncation mutant 866-1206
CC (numbered relative to the full length NS2/3 protein)
XX
SQ Sequence 341 AA;

Query Match 100.0%; Score 1589; DB 5; Length 341;
Best Local Similarity 100.0%; Pred. No. 6e-146;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy           1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          39 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 98

Qy          61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          99 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 158

Qy         121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         159 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 218

Qy         181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         219 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 278

Qy         241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         279 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 338

Qy         301 TMR 303
               |||
Db         339 TMR 341
```

RESULT 4
ABG32186
ID ABG32186 standard; protein; 352 AA.
XX
AC ABG32186;
XX
DT 05-NOV-2002 (first entry)
XX
DE HCV protease NS2/3 truncation mutant 855-1206.
XX
KW HCV; enzyme; protease; NS2/3 (855-1206); hepatitis C virus infection;
KW chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW chaotropic agent; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO200248375-A2.
XX
PD 20-JUN-2002.

XX
PF 13-DEC-2001; 2001WO-CA001796.
XX
PR 15-DEC-2000; 2000US-0256031P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
XX
DR WPI; 2002-599511/64.
XX
PT Novel polypeptide for screening inhibitors of non-structural proteases
PT useful as therapeutic agents against hepatitis C virus, comprises full
PT length non-structural protease, or its truncation.
XX
PS Claim 41; Page 61-62; 67pp; English.
XX
CC The invention relates to an isolated polypeptide consisting of a full-
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC residue amino acid 810 to 906, or having a minimal amino acid sequence
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC NS2/3 protease. Also included are (1) a composition (C) comprising an
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC its truncation or a mutated sequence, where the protease is in a solution
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC protease, involving isolating the protease in the presence of a
CC chaotropic agent, refolding the isolated protease by contacting it with a
CC reducing agent, and LDAO in the presence of reduced concentration of the
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC protease, involving diluting refolded inactive NS2/3 protease in a medium
CC containing an activation detergent to induce auto-cleavage of the NS2/3
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC protease, involving incubating the active NS2/3 protease produced by M2
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC cleavage products or their fragments, and measuring the presence or
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC active NS2/3 protease, involving carrying out M3 in the presence of, or
CC absence of the potential inhibitor, comparing the amount of uncleaved
CC NS2/3 protease, cleavage products or their fragments. The protease is
CC useful for detailed biochemical characterisation of the enzymes and in
CC the development of in vitro assays for screening novel inhibitors of
CC NS2/3 protease which are useful as therapeutic agents against HCV
CC infection (which causes chronic liver disease, cirrhosis and end-stage
CC liver disease. M1 is useful for high level production of protease. The
CC present sequence represents the NS2/3 truncation mutant 855-1206
CC (numbered relative to the full length NS2/3 protein)
XX
SQ Sequence 352 AA;

Query Match 100.0%; Score 1589; DB 5; Length 352;
Best Local Similarity 100.0%; Pred. No. 6.3e-146;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         50 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 109

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        110 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 169

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        170 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 229

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        230 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 289

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        290 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 349

Qy        301 TMR 303
              |||
Db        350 TMR 352
```

RESULT 5
ABG32185
ID ABG32185 standard; protein; 380 AA.
XX
AC ABG32185;
XX
DT 05-NOV-2002 (first entry)
XX
DE HCV protease NS2/3 truncation mutant 827-1206.
XX
KW HCV; enzyme; protease; NS2/3 (827-1206); hepatitis C virus infection;
KW chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW chaotropic agent; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO200248375-A2.
XX
PD 20-JUN-2002.
XX
PF 13-DEC-2001; 2001WO-CA001796.
XX
PR 15-DEC-2000; 2000US-0256031P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
XX
DR WPI; 2002-599511/64.
XX
PT Novel polypeptide for screening inhibitors of non-structural proteases
PT useful as therapeutic agents against hepatitis C virus, comprises full
PT length non-structural protease, or its truncation.
XX
PS Claim 41; Page 60-61; 67pp; English.
XX
CC The invention relates to an isolated polypeptide consisting of a full-
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC residue amino acid 810 to 906, or having a minimal amino acid sequence
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC NS2/3 protease. Also included are (1) a composition (C) comprising an
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC its truncation or a mutated sequence, where the protease is in a solution
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC protease, involving isolating the protease in the presence of a
CC chaotropic agent, refolding the isolated protease by contacting it with a
CC reducing agent, and LDAO in the presence of reduced concentration of the
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC protease, involving diluting refolded inactive NS2/3 protease in a medium
CC containing an activation detergent to induce auto-cleavage of the NS2/3
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC protease, involving incubating the active NS2/3 protease produced by M2
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC cleavage products or their fragments, and measuring the presence or
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC active NS2/3 protease, involving carrying out M3 in the presence of, or
CC absence of the potential inhibitor, comparing the amount of uncleaved
CC NS2/3 protease, cleavage products or their fragments. The protease is

```
CC   useful for detailed biochemical characterisation of the enzymes and in
CC   the development of in vitro assays for screening novel inhibitors of
CC   NS2/3 protease which are useful as therapeutic agents against HCV
CC   infection (which causes chronic liver disease, cirrhosis and end-stage
CC   liver disease. M1 is useful for high level production of protease. The
CC   present sequence represents the NS2/3 truncation mutant 827-1206
CC   (numbered relative to the full length NS2/3 protein)
XX
SQ   Sequence 380 AA;

  Query Match            100.0%;  Score 1589;  DB 5;  Length 380;
  Best Local Similarity  100.0%;  Pred. No. 7e-146;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 137

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 197

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 257

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 317

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 377

Qy        301 TMR 303
              |||
Db        378 TMR 380


RESULT 6
ABG32184
ID   ABG32184 standard; protein; 393 AA.
XX
AC   ABG32184;
XX
DT   05-NOV-2002  (first entry)
XX
DE   HCV protease NS2/3 truncation mutant 815-1206.
XX
KW   HCV; enzyme; protease; NS2/3 (815-1206); hepatitis C virus infection;
KW   chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW   hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW   chaotropic agent; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
PN   WO200248375-A2.
XX
PD   20-JUN-2002.
XX
PF   13-DEC-2001; 2001WO-CA001796.
XX
PR   15-DEC-2000; 2000US-0256031P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Thibeault D,  Lamarre D,  Maurice R,  Pilote L,  Pause A;
XX
DR   WPI; 2002-599511/64.
XX
PT   Novel polypeptide for screening inhibitors of non-structural proteases
PT   useful as therapeutic agents against hepatitis C virus, comprises full
PT   length non-structural protease, or its truncation.
XX
PS   Claim 41; Page 59-60; 67pp; English.
XX
CC   The invention relates to an isolated polypeptide consisting of a full-
CC   length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC   to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC   residue amino acid 810 to 906, or having a minimal amino acid sequence
CC   from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC   NS2/3 protease. Also included are (1) a composition (C) comprising an
CC   isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC   its truncation or a mutated sequence, where the protease is in a solution
CC   comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC   to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC   appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC   protease, involving isolating the protease in the presence of a
CC   chaotropic agent, refolding the isolated protease by contacting it with a
CC   reducing agent, and LDAO in the presence of reduced concentration of the
CC   chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC   protease, involving diluting refolded inactive NS2/3 protease in a medium
CC   containing an activation detergent to induce auto-cleavage of the NS2/3
CC   protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC   protease, involving incubating the active NS2/3 protease produced by M2
CC   for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC   cleavage products or their fragments, and measuring the presence or
CC   absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC   ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC   active NS2/3 protease, involving carrying out M3 in the presence of, or
CC   absence of the potential inhibitor, comparing the amount of uncleaved
CC   NS2/3 protease, cleavage products or their fragments. The protease is
CC   useful for detailed biochemical characterisation of the enzymes and in
CC   the development of in vitro assays for screening novel inhibitors of
CC   NS2/3 protease which are useful as therapeutic agents against HCV
CC   infection (which causes chronic liver disease, cirrhosis and end-stage
CC   liver disease. M1 is useful for high level production of protease. The
CC   present sequence represents the NS2/3 truncation mutant 815-1206
CC   (numbered relative to the full length NS2/3 protein)
XX
SQ   Sequence 393 AA;

  Query Match            100.0%;  Score 1589;  DB 5;  Length 393;
  Best Local Similarity  100.0%;  Pred. No. 7.3e-146;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         91 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 150

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        151 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 210

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 270

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        271 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 330

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        331 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 390

Qy        301 TMR 303
              |||
Db        391 TMR 393


RESULT 7
ABG32181
ID   ABG32181 standard; protein; 409 AA.
XX
```

```
AC   ABG32181;
XX
DT   05-NOV-2002  (first entry)
XX
DE   HCV protease NS2/3 (810-1206).
XX
KW   HCV; enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;
KW   chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW   hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW   chaotropic agent; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         398. .409
FT                   /note= "Streptavidin tag"
XX
PN   WO200248375-A2.
XX
PD   20-JUN-2002.
XX
PF   13-DEC-2001; 2001WO-CA001796.
XX
PR   15-DEC-2000; 2000US-0256031P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Thibeault D,  Lamarre D,  Maurice R,  Pilote L,  Pause A;
XX
DR   WPI; 2002-599511/64.
DR   N-PSDB; ABK90406.
XX
PT   Novel polypeptide for screening inhibitors of non-structural proteases
PT   useful as therapeutic agents against hepatitis C virus, comprises full
PT   length non-structural protease, or its truncation.
XX
PS   Claim 42; Fig 1B; 67pp; English.
XX
CC   The invention relates to an isolated polypeptide consisting of a full-
CC   length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC   to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC   residue amino acid 810 to 906, or having a minimal amino acid sequence
CC   from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC   NS2/3 protease. Also included are (1) a composition (C) comprising an
CC   isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC   its truncation or a mutated sequence, where the protease is in a solution
CC   comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC   to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC   appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC   protease, involving isolating the protease in the presence of a
CC   chaotropic agent, refolding the isolated protease by contacting it with a
CC   reducing agent, and LDAO in the presence of reduced concentration of the
CC   chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC   protease, involving diluting refolded inactive NS2/3 protease in a medium
CC   containing an activation detergent to induce auto-cleavage of the NS2/3
CC   protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC   protease, involving incubating the active NS2/3 protease produced by M2
CC   for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC   cleavage products or their fragments, and measuring the presence or
CC   absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC   ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC   active NS2/3 protease, involving carrying out M3 in the presence of, or
CC   absence of the potential inhibitor, comparing the amount of uncleaved
CC   NS2/3 protease, cleavage products or their fragments. The protease is
CC   useful for detailed biochemical characterisation of the enzymes and in
CC   the development of in vitro assays for screening novel inhibitors of
CC   NS2/3 protease which are useful as therapeutic agents against HCV
CC   infection (which causes chronic liver disease, cirrhosis and end-stage
CC   liver disease. M1 is useful for high level production of protease. The
CC   present sequence represents the NS2/3 (810-1206) protein, which has a C-
CC   terminal streptavividin tag
XX
SQ   Sequence 409 AA;

  Query Match             100.0%;  Score 1589;  DB 5;  Length 409;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-146;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         95 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 154

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        155 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 214

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        215 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 274

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        275 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 334

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        335 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 394

Qy        301 TMR 303
              |||
Db        395 TMR 397


RESULT 8
ABG32191
ID   ABG32191 standard; protein; 303 AA.
XX
AC   ABG32191;
XX
DT   05-NOV-2002  (first entry)
XX
DE   HCV protease NS2/3 truncation 904-1206/Cys993Ala.
XX
KW   HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;
KW   chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW   hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW   chaotropic agent; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 90
FT                   /note= "Wild-type Cys substituted by Ala"
XX
PN   WO200248375-A2.
XX
PD   20-JUN-2002.
XX
PF   13-DEC-2001; 2001WO-CA001796.
XX
PR   15-DEC-2000; 2000US-0256031P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Thibeault D,  Lamarre D,  Maurice R,  Pilote L,  Pause A;
XX
DR   WPI; 2002-599511/64.
XX
PT   Novel polypeptide for screening inhibitors of non-structural proteases
PT   useful as therapeutic agents against hepatitis C virus, comprises full
PT   length non-structural protease, or its truncation.
XX
PS   Disclosure; Page 65-66; 67pp; English.
XX
```

CC The invention relates to an isolated polypeptide consisting of a full-
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC residue amino acid 810 to 906, or having a minimal amino acid sequence
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC NS2/3 protease. Also included are (1) a composition (C) comprising an
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC its truncation or a mutated sequence, where the protease is in a solution
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC protease, involving isolating the protease in the presence of a
CC chaotropic agent, refolding the isolated protease by contacting it with a
CC reducing agent, and LDAO in the presence of reduced concentration of the
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC protease, involving diluting refolded inactive NS2/3 protease in a medium
CC containing an activation detergent to induce auto-cleavage of the NS2/3
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC protease, involving incubating the active NS2/3 protease produced by M2
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC cleavage products or their fragments, and measuring the presence or
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC active NS2/3 protease, involving carrying out M3 in the presence of, or
CC absence of the potential inhibitor, comparing the amount of uncleaved
CC NS2/3 protease, cleavage products or their fragments. The protease is
CC useful for detailed biochemical characterisation of the enzymes and in
CC the development of in vitro assays for screening novel inhibitors of
CC NS2/3 protease which are useful as therapeutic agents against HCV
CC infection (which causes chronic liver disease, cirrhosis and end-stage
CC liver disease. M1 is useful for high level production of protease. The
CC present sequence represents the NS2/3 truncation 904-1206 mutant
CC Cys993Ala (numbered relative to the full length NS2/3 protein) a mutant
CC devoid of autocatalytic activity
XX
SQ Sequence 303 AA;

Query Match 99.4%; Score 1580; DB 5; Length 303;
Best Local Similarity 99.7%; Pred. No. 3.8e-145;
Matches 302; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db         61 LRDLAVAVEPVIFSDMEVKIITWGADTAAAGDIISGLPVSARRGREILLGPADNFEGQGW 120

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300

Qy        301 TMR 303
              |||
Db        301 TMR 303
```

RESULT 9
ABG32189
ID ABG32189 standard; protein; 303 AA.
XX
AC ABG32189;
XX
DT 05-NOV-2002 (first entry)
XX
DE HCV protease NS2/3 truncation 904-1206/His952Ala.
XX
KW HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;
KW chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW chaotropic agent; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 49
FT /note= "Wild-type His substituted by Ala"
XX
PN WO200248375-A2.
XX
PD 20-JUN-2002.
XX
PF 13-DEC-2001; 2001WO-CA001796.
XX
PR 15-DEC-2000; 2000US-0256031P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
XX
DR WPI; 2002-599511/64.
XX
PT Novel polypeptide for screening inhibitors of non-structural proteases
PT useful as therapeutic agents against hepatitis C virus, comprises full
PT length non-structural protease, or its truncation.
XX
PS Example 7; Fig 8; 67pp; English.
XX
CC The invention relates to an isolated polypeptide consisting of a full-
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC residue amino acid 810 to 906, or having a minimal amino acid sequence
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC NS2/3 protease. Also included are (1) a composition (C) comprising an
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC its truncation or a mutated sequence, where the protease is in a solution
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC protease, involving isolating the protease in the presence of a
CC chaotropic agent, refolding the isolated protease by contacting it with a
CC reducing agent, and LDAO in the presence of reduced concentration of the
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC protease, involving diluting refolded inactive NS2/3 protease in a medium
CC containing an activation detergent to induce auto-cleavage of the NS2/3
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC protease, involving incubating the active NS2/3 protease produced by M2
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC cleavage products or their fragments, and measuring the presence or
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC active NS2/3 protease, involving carrying out M3 in the presence of, or
CC absence of the potential inhibitor, comparing the amount of uncleaved
CC NS2/3 protease, cleavage products or their fragments. The protease is
CC useful for detailed biochemical characterisation of the enzymes and in
CC the development of in vitro assays for screening novel inhibitors of
CC NS2/3 protease which are useful as therapeutic agents against HCV
CC infection (which causes chronic liver disease, cirrhosis and end-stage
CC liver disease. M1 is useful for high level production of protease. The
CC present sequence represents the NS2/3 truncation 904-1206 mutant
CC His952Ala (numbered relative to the full length NS2/3 protein) a mutant
CC devoid of autocatalytic activity
XX
SQ Sequence 303 AA;

Query Match 99.4%; Score 1579; DB 5; Length 303;

```
  Best Local Similarity   99.7%;  Pred. No. 4.8e-145;
  Matches  302;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDALTPLQDWAHAG 60

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300

Qy        301 TMR 303
              |||
Db        301 TMR 303
```

```
RESULT 10
ABG32190
ID   ABG32190 standard; protein; 301 AA.
XX
AC   ABG32190;
XX
DT   05-NOV-2002  (first entry)
XX
DE   HCV protease NS2/3 truncation 904-1206/deltaLeu1026-Ala1027.
XX
KW   HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;
KW   chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW   hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW   chaotropic agent; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 122. .123
FT                   /note= "Wild-type Leu-Leu-Ala-Pro substituted by Leu-Pro"
XX
PN   WO200248375-A2.
XX
PD   20-JUN-2002.
XX
PF   13-DEC-2001; 2001WO-CA001796.
XX
PR   15-DEC-2000; 2000US-0256031P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Thibeault D,  Lamarre D,  Maurice R,  Pilote L,  Pause A;
XX
DR   WPI; 2002-599511/64.
XX
PT   Novel polypeptide for screening inhibitors of non-structural proteases
PT   useful as therapeutic agents against hepatitis C virus, comprises full
PT   length non-structural protease, or its truncation.
XX
PS   Example 7; Page 64-65; 67pp; English.
XX
CC   The invention relates to an isolated polypeptide consisting of a full-
CC   length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC   to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC   residue amino acid 810 to 906, or having a minimal amino acid sequence
CC   from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC   NS2/3 protease. Also included are (1) a composition (C) comprising an
CC   isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC   its truncation or a mutated sequence, where the protease is in a solution
CC   comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC   to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC   appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC   protease, involving isolating the protease in the presence of a
CC   chaotropic agent, refolding the isolated protease by contacting it with a
CC   reducing agent, and LDAO in the presence of reduced concentration of the
CC   chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC   protease, involving diluting refolded inactive NS2/3 protease in a medium
CC   containing an activation detergent to induce auto-cleavage of the NS2/3
CC   protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC   protease, involving incubating the active NS2/3 protease produced by M2
CC   for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC   cleavage products or their fragments, and measuring the presence or
CC   absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC   ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC   active NS2/3 protease, involving carrying out M3 in the presence of, or
CC   absence of the potential inhibitor, comparing the amount of uncleaved
CC   NS2/3 protease, cleavage products or their fragments. The protease is
CC   useful for detailed biochemical characterisation of the enzymes and in
CC   the development of in vitro assays for screening novel inhibitors of
CC   NS2/3 protease which are useful as therapeutic agents against HCV
CC   infection (which causes chronic liver disease, cirrhosis and end-stage
CC   liver disease. M1 is useful for high level production of protease. The
CC   present sequence represents the NS2/3 truncation 904-1206 mutant
CC   deltaLeu1026-Ala1027 (numbered relative to the full length NS2/3 protein)
CC   a mutant devoid of autocatalytic activity
XX
SQ   Sequence 301 AA;
```

```
  Query Match             98.8%;  Score 1570;  DB 5;  Length 301;
  Best Local Similarity   99.3%;  Pred. No. 3.6e-144;
  Matches  301;  Conservative    0;  Mismatches    0;  Indels    2;  Gaps    1;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RL--PITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 178

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 238

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 298

Qy        301 TMR 303
              |||
Db        299 TMR 301
```

```
RESULT 11
ABG32188
ID   ABG32188 standard; protein; 292 AA.
XX
AC   ABG32188;
XX
DT   05-NOV-2002  (first entry)
XX
DE   HCV protease NS2/3 truncation mutant 915-1206.
XX
```

```
KW   HCV; enzyme; protease; NS2/3 (915-1206); hepatitis C virus infection;
KW   chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW   hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW   chaotropic agent; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
PN   WO200248375-A2.
XX
PD   20-JUN-2002.
XX
PF   13-DEC-2001; 2001WO-CA001796.
XX
PR   15-DEC-2000; 2000US-0256031P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Thibeault D,  Lamarre D,  Maurice R,  Pilote L,  Pause A;
XX
DR   WPI; 2002-599511/64.
XX
PT   Novel polypeptide for screening inhibitors of non-structural proteases
PT   useful as therapeutic agents against hepatitis C virus, comprises full
PT   length non-structural protease, or its truncation.
XX
PS   Claim 41; Page 63; 67pp; English.
XX
CC   The invention relates to an isolated polypeptide consisting of a full-
CC   length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC   to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC   residue amino acid 810 to 906, or having a minimal amino acid sequence
CC   from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC   NS2/3 protease. Also included are (1) a composition (C) comprising an
CC   isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC   its truncation or a mutated sequence, where the protease is in a solution
CC   comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC   to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC   appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC   protease, involving isolating the protease in the presence of a
CC   chaotropic agent, refolding the isolated protease by contacting it with a
CC   reducing agent, and LDAO in the presence of reduced concentration of the
CC   chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC   protease, involving diluting refolded inactive NS2/3 protease in a medium
CC   containing an activation detergent to induce auto-cleavage of the NS2/3
CC   protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC   protease, involving incubating the active NS2/3 protease produced by M2
CC   for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC   cleavage products or their fragments, and measuring the presence or
CC   absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC   ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC   active NS2/3 protease, involving carrying out M3 in the presence of, or
CC   absence of the potential inhibitor, comparing the amount of uncleaved
CC   NS2/3 protease, cleavage products or their fragments. The protease is
CC   useful for detailed biochemical characterisation of the enzymes and in
CC   the development of in vitro assays for screening novel inhibitors of
CC   NS2/3 protease which are useful as therapeutic agents against HCV
CC   infection (which causes chronic liver disease, cirrhosis and end-stage
CC   liver disease. M1 is useful for high level production of protease. The
CC   present sequence represents the NS2/3 truncation mutant 915-1206
CC   (numbered relative to the full length NS2/3 protein)
XX
SQ   Sequence 292 AA;

  Query Match             96.4%;  Score 1532;  DB 5;  Length 292;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-140;
  Matches  292;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         12 AQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPV 71
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPV 60

Qy         72 IFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQ 131
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQ 120

Qy        132 QTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPK 191
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPK 180

Qy        192 GPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLS 251
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLS 240

Qy        252 PRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 292
```

```
RESULT 12
ABG30601
ID   ABG30601 standard; protein; 2201 AA.
XX
AC   ABG30601;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #10.
XX
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 882
FT                   /label= Arg, Lys
FT   Misc-difference 2183
FT                   /note= "Wild type Met substituted by Thr"
XX
PN   WO200252015-A2.
XX
PD   04-JUL-2002.
XX
PF   20-DEC-2001; 2001WO-CA001843.
XX
PR   22-DEC-2000; 2000US-0257857P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Kukolj G,  Pause A;
XX
DR   WPI; 2002-575382/61.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
XX
PS   Claim 3; Page; 140pp; English.
XX
CC   The invention describes a self-replicating hepatitis C virus (HCV)
CC   polynucleotide molecule comprising a 5'-non translated region (NTR),
CC   where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC   region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC   replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC   potential inhibitors of HCV replication. The HCV RNA molecule is also
CC   useful for efficiently establishing cell culture replication. The self-
CC   replicating polynucleotide molecule contains a 5'-NTR, where G at
CC   position 1 is substituted for A, and therefore provides an alternative to
CC   existing systems comprising a self-replicating HCV RNA molecule that, in
CC   conjunction with mutations in the HCV non-structural region, such as the
CC   G(2042)C/R mutations, transduces and/or replicates with greater
CC   efficiency. This amino acid sequence represents a mutant of the hepatitis
CC   C virus replicon APGK12 and contains the viral protease NS2/3, protease
CC   complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
```

```
CC   This sequence does not appear in the specification but has been created
CC   from the wild type sequence shown in ABG30580 using information given in
CC   the claims of the invention
XX
SQ   Sequence 2201 AA;

  Query Match             96.3%;  Score 1530;  DB 5;  Length 2201;
  Best Local Similarity   95.0%;  Pred. No. 4.6e-139;
  Matches  288;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||| |||||||||| |||||||| |||||||||||||||||:||||||
Db         95 AGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAG 154

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |:||||||||||||| |||||||||| |||||: |||||
Db        155 LRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQGW 214

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||:||||||||||||||||||||||||:|
Db        215 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 274

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db        275 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 334

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        335 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 394

Qy        301 TMR 303
              |||
Db        395 TMR 397


RESULT 13
ABG30591
ID   ABG30591 standard; protein; 2201 AA.
XX
AC   ABG30591;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #3.
XX
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 751
FT                   /note= "Wild type Ser substituted by Gly"
FT   Misc-difference 882
FT                   /label= Arg, Lys
XX
PN   WO200252015-A2.
XX
PD   04-JUL-2002.
XX
PF   20-DEC-2001; 2001WO-CA001843.
XX
PR   22-DEC-2000; 2000US-0257857P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Kukolj G,  Pause A;
XX
DR   WPI; 2002-575382/61.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
XX
PS   Claim 3; Page; 140pp; English.
XX
CC   The invention describes a self-replicating hepatitis C virus (HCV)
CC   polynucleotide molecule comprising a 5'-non translated region (NTR),
CC   where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC   region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC   replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC   potential inhibitors of HCV replication. The HCV RNA molecule is also
CC   useful for efficiently establishing cell culture replication. The self-
CC   replicating polynucleotide molecule contains a 5'-NTR, where G at
CC   position 1 is substituted for A, and therefore provides an alternative to
CC   existing systems comprising a self-replicating HCV RNA molecule that, in
CC   conjunction with mutations in the HCV non-structural region, such as the
CC   G(2042)C/R mutations, transduces and/or replicates with greater
CC   efficiency. This amino acid sequence represents a mutant of the hepatitis
CC   C virus replicon APGK12 and contains the viral protease NS2/3, protease
CC   complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
CC   This sequence does not appear in the specification but has been created
CC   from the wild type sequence shown in ABG30580 using information given in
CC   the claims of the invention
XX
SQ   Sequence 2201 AA;

  Query Match             96.3%;  Score 1530;  DB 5;  Length 2201;
  Best Local Similarity   95.0%;  Pred. No. 4.6e-139;
  Matches  288;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||| |||||||||| |||||||| |||||||||||||||||:||||||
Db         95 AGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAG 154

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |:||||||||||||| |||||||||| |||||: |||||
Db        155 LRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQGW 214

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||:||||||||||||||||||||||||:|
Db        215 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 274

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db        275 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 334

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        335 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 394

Qy        301 TMR 303
              |||
Db        395 TMR 397


RESULT 14
ABG30600
ID   ABG30600 standard; protein; 2201 AA.
XX
AC   ABG30600;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #9.
XX
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 882
FT                   /label= Arg, Lys
```

```
FT   Misc-difference 1357
FT                   /note= "Wild type Pro substituted by Leu"
XX
PN   WO200252015-A2.
XX
PD   04-JUL-2002.
XX
PF   20-DEC-2001; 2001WO-CA001843.
XX
PR   22-DEC-2000; 2000US-0257857P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Kukolj G,  Pause A;
XX
DR   WPI; 2002-575382/61.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
XX
PS   Claim 3; Page; 140pp; English.
XX
CC   The invention describes a self-replicating hepatitis C virus (HCV)
CC   polynucleotide molecule comprising a 5'-non translated region (NTR),
CC   where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC   region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC   replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC   potential inhibitors of HCV replication. The HCV RNA molecule is also
CC   useful for efficiently establishing cell culture replication. The self-
CC   replicating polynucleotide molecule contains a 5'-NTR, where G at
CC   position 1 is substituted for A, and therefore provides an alternative to
CC   existing systems comprising a self-replicating HCV RNA molecule that, in
CC   conjunction with mutations in the HCV non-structural region, such as the
CC   G(2042)C/R mutations, transduces and/or replicates with greater
CC   efficiency. This amino acid sequence represents a mutant of the hepatitis
CC   C virus replicon APGK12 and contains the viral protease NS2/3, protease
CC   complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
CC   This sequence does not appear in the specification but has been created
CC   from the wild type sequence shown in ABG30580 using information given in
CC   the claims of the invention
XX
SQ   Sequence 2201 AA;

  Query Match             96.3%;  Score 1530;  DB 5;  Length 2201;
  Best Local Similarity   95.0%;  Pred. No. 4.6e-139;
  Matches  288;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||| |||||||||| |||||||| ||||||||||||||||||:||||||
Db         95 AGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAG 154

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:||||| |:||||||||||||| |||||||||| ||||||: |||||
Db        155 LRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQGW 214

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||:||||||||||||||||||||||||||:|
Db        215 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 274

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
Db        275 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 334

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        335 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 394

Qy        301 TMR 303
              |||
Db        395 TMR 397
```

```
RESULT 15
ABG30581
ID   ABG30581 standard; protein; 2201 AA.
XX
AC   ABG30581;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #1.
XX
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; NS2/3; NS3/4; NS3; NS5B.
XX
OS   Hepatitis C virus.
XX
PN   WO200252015-A2.
XX
PD   04-JUL-2002.
XX
PF   20-DEC-2001; 2001WO-CA001843.
XX
PR   22-DEC-2000; 2000US-0257857P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Kukolj G,  Pause A;
XX
DR   WPI; 2002-575382/61.
DR   N-PSDB; ABK88573.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
XX
PS   Disclosure; Page 49-58; 140pp; English.
XX
CC   The invention describes a self-replicating hepatitis C virus (HCV)
CC   polynucleotide molecule comprising a 5'-non translated region (NTR),
CC   where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC   region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC   replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC   potential inhibitors of HCV replication. The HCV RNA molecule is also
CC   useful for efficiently establishing cell culture replication. The self-
CC   replicating polynucleotide molecule contains a 5'-NTR, where G at
CC   position 1 is substituted for A, and therefore provides an alternative to
CC   existing systems comprising a self-replicating HCV RNA molecule that, in
CC   conjunction with mutations in the HCV non-structural region, such as the
CC   G(2042)C/R mutations, transduces and/or replicates with greater
CC   efficiency. This amino acid sequence is encoded by the hepatitis C virus
CC   replicon APGK12 and contains the viral protease NS2/3, protease complex
CC   NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B
XX
SQ   Sequence 2201 AA;

  Query Match             96.3%;  Score 1530;  DB 5;  Length 2201;
  Best Local Similarity   95.0%;  Pred. No. 4.6e-139;
  Matches  288;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||| |||||||||| |||||||| ||||||||||||||||||:||||||
Db         95 AGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAG 154

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:||||| |:||||||||||||| |||||||||| ||||||: |||||
Db        155 LRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQGW 214

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||:||||||||||||||||||||||||||:|
Db        215 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 274

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
Db        275 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 334
```

```
Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        335 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 394

Qy        301 TMR 303
              |||
Db        395 TMR 397

Search completed: May  6, 2004, 09:30:44
Job time : 42.2434 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:22:36 ; Search time 9.86806 Seconds
(without alignments)
2953.573 Million cell updates/sec

Title: US-10-650-585-10
Perfect score: 1589
Sequence: 1 AGITKVPYFVRAQGLIRACM.....RGVAKAVDFIPVESMETTMR 303

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1532	96.4	3010	1	A45573	genome polyprotein
2	1527	96.1	3010	1	GNWVCJ	genome polyprotein
3	1514	95.3	3010	1	GNWVTW	genome polyprotein
4	1486	93.5	3010	1	S18030	genome polyprotein
5	1478	93.0	3010	1	GNWVTC	genome polyprotein
6	1403	88.3	3011	1	S40770	genome polyprotein
7	1398	88.0	3011	1	GNWVC3	genome polyprotein
8	1385	87.2	3011	1	GNWVCH	genome polyprotein
9	1234	77.7	3014	1	JC5620	genome polyprotein
10	1172	73.8	3033	1	JQ1303	genome polyprotein
11	1150	72.4	3033	1	GNWVJ8	genome polyprotein
12	397.5	25.0	3005	2	T08841	polyprotein - dour
13	341	21.5	2970	2	T08839	polyprotein - marm
14	101	6.4	600	2	B46642	DNA-directed DNA p
15	99.5	6.3	353	2	G87392	conserved hypothet
16	97.5	6.1	1085	2	T03531	cobN protein homol
17	95.5	6.0	470	2	JC4098	tetracycline 6-hyd
18	93	5.9	660	1	VHWWH2	structural protein
19	92.5	5.8	706	2	S33761	transferrin precur
20	92.5	5.8	716	2	G83612	hypothetical prote
21	91	5.7	904	2	A84212	hypothetical prote
22	90.5	5.7	868	2	H81775	aconitate hydratas
23	90	5.7	2796	2	JC4743	fatty-acid synthas
24	90	5.7	7463	2	T36248	CDA peptide synthe
25	89.5	5.6	659	1	B44212	structural protein
26	88	5.5	3414	1	GNWVNE	genome polyprotein
27	87	5.5	3412	1	GNWVTB	genome polyprotein
28	86.5	5.4	470	1	NMIVW8	exo-alpha-sialidas
29	85.5	5.4	348	2	H70549	probable pdhB prot
30	85	5.3	470	1	NMIV9	exo-alpha-sialidas
31	85	5.3	707	2	D84154	cadmium-transporti
32	84.5	5.3	347	2	S44167	malate dehydrogena
33	84.5	5.3	1057	2	I38171	hugl protein - hum
34	84	5.3	233	2	T35594	hypothetical prote
35	84	5.3	3069	2	H70656	fatty-acid synthas
36	83.5	5.3	299	2	AH3447	cytochrome-c oxida
37	83.5	5.3	315	2	AG2361	hypothetical prote
38	83.5	5.3	538	2	S22409	D-alanyl-D-alanine
39	83.5	5.3	990	2	S67499	glutamate synthase
40	83	5.2	398	2	B71284	probable periplasm
41	82.5	5.2	428	2	AF0241	probable coenzyme
42	82.5	5.2	453	1	NMIV3	exo-alpha-sialidas
43	82.5	5.2	868	2	C81200	aconitate hydratas
44	82.5	5.2	5627	2	C83339	hypothetical prote
45	82	5.2	404	2	F87393	aminotransferase,

ALIGNMENTS

RESULT 1
A45573
genome polyprotein - hepatitis C virus (strain JT)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C;Accession: A45573
R;Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,
Virus Res. 23, 39-53, 1992
A;Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: s
A;Reference number: A45573; MUID:92295714; PMID:1318627
A;Accession: A45573
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-3010 <TAN>
A;Cross-references: GB:D11168; GB:D01171; NID:g221612; PIDN:BAA01943.1; PID:g221613
A;Experimental source: HCV-JT
A;Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBIP:106207)
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 96.4%; Score 1532; DB 1; Length 3010;
Best Local Similarity 95.4%; Pred. No. 7.3e-124;
Matches 289; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

```
Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              | || :||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db        904 AAITAMPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:||||| |||||||||||||| |||||||||||||||||: |||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGPADSIEGQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              |||||||||:||||||||||:|||||||||||||||||||||||||||||||||||||||
Db       1024 RLLAPITAYAQQTRGLLGCIVTSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              |||||||||||||||||||||||||||| |||||||:||||||||||||||||||||||
```

```
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWHAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 RRGDGRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 2
GNWVCJ
genome polyprotein - hepatitis C virus (strain J)
N;Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructura
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001
C;Accession: A39253; PS0086
R;Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimot
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
A;Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients
A;Reference number: A39253; MUID:91088550; PMID:2175903
A;Accession: A39253
A;Molecule type: genomic RNA
A;Residues: 1-3010 <KAT>
A;Cross-references: GB:D90208; NID:g221610; PIDN:BAA14233.1; PID:g221611
R;Kato, N.; Ohkoshi, S.; Shimotohno, K.
Proc. Jpn. Acad. 65B, 219-223, 1989
A;Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari
A;Reference number: PS0085
A;Accession: PS0086
A;Molecule type: genomic RNA
A;Residues: 2650-2707 <KA2>
A;Experimental source: Japanese isolate
C;Comment: The cleavage sites of this polyprotein have not been determined.
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2

  Query Match             96.1%;  Score 1527;  DB 1;  Length 3010;
  Best Local Similarity   94.4%;  Pred. No. 2e-123;
  Matches  286;  Conservative   11;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||:|||||||||||||||||| |||||||||||||||||||||||||||||:||||||
Db        904 AGITRVPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLRDWAHAG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |:|||||||||||||||||||||:|||||||:|  |||
Db        964 LRDLAVAVEPVVFSDMETKLITWGADTAACGDIISGLPVSARRGKEILLGPADSFGEQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||:|||||:||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVDGEVQVLSTATQSFLATCVNGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||:|||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWPAPPGARSMTPCTCGSSDLYLVTRHADVVPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              |||||||||||||:|||||||||||||||||||||||| ||||||||||||||||||||||
Db       1144 RRGDSRGSLLSPRPISYLKGSSGGPLLCPSGHVVGIFRAAVCTRGVAKAVDFIPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 3
GNWVTW
genome polyprotein - hepatitis C virus (strain Taiwan)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructur
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: A40244
R;Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992
A;Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the 5
A;Reference number: A40244; MUID:92230206; PMID:1314449
A;Accession: A40244
A;Molecule type: genomic RNA
A;Residues: 1-3010 <CHE>
A;Cross-references: GB:M84754
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

  Query Match             95.3%;  Score 1514;  DB 1;  Length 3010;
  Best Local Similarity   93.1%;  Pred. No. 2.6e-122;
  Matches  282;  Conservative   12;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||:|::|||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        904 AGLTRIPYFVRAQGLIRACMLVRKAAGGHYVQMALMKLAALTGTYVYDHLTPLQDWAHTG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |||||||||||||| ||||||||||||||||||: ||:||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGPADSLEGRGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              |||||||||:|||||| ||||||||||||||||||||||||||||:|||||||:|
Db       1024 RLLAPITAYAQQTRGLFGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCINGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              |||||||||||||||||||||||||||| || ||||:|||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWHAPQGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              |||||||||||||:|||||||||||||||||||||||| ||||||||||||||||||||||
Db       1144 RRGDSRGSLLSPRPISYLKGSSGGPLLCPSGHVVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 4
S18030
genome polyprotein - hepatitis C virus (isolate JK1)
```

```
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Variety: isolate JK1
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001
C;Accession: S18030; S33570; A48332; S18029
R;Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A;Description: A whole genome of hepatitis C virus cDNA was isolated from a single patie
A;Reference number: S18028
A;Accession: S18030
A;Molecule type: genomic RNA
A;Residues: 1-3010 <HON>
A;Cross-references: EMBL:X61596; NID:g59478; PIDN:CAA43793.1; PID:g59479
A;Experimental source: isolate JK1 from an individual
R;Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.
Arch. Virol. 128, 163-169, 1993
A;Title: Sequence analysis of putative structural regions of hepatitis C virus isolated
A;Reference number: A48332; MUID:93119270; PMID:8380322
A;Accession: S33570
A;Molecule type: genomic RNA
A;Residues: 1-547,'T',549-621,'V',623-624,'S',626-652,'DL',655-761,'T',763-782 <HOW>
A;Cross-references: EMBL:X61591
A;Note: this sequence is inconsistent with the nucleotide translation
A;Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320
 as Trp, and TTC for residue 771 as Ser
A;Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBIP:121748)
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (As

  Query Match           93.5%;  Score 1486;  DB 1;  Length 3010;
  Best Local Similarity   92.4%;  Pred. No. 7e-120;
  Matches  280;  Conservative    9;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||:|:|||||||||||:|||||||| |||||||| |||||||||||||||||:||||:|
Db        904 AGLTRVPYFVRAQGLLRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHSG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||| |||||:||||| |||||||||||||| |||||||||||||||| |   |||
Db        964 LRDLVVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGPADGFREQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              |||||||||||||||| |||:||||||||||||| |||||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLFGCIVTSLTGRDKNQVEGEAQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||| |||||||||||||||  ||  |:|||| ||||||||||||||||
Db       1084 GAGSKTLAGPKGPINQMYTNVDQDLVGWQAPSGAASLTPCTYGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 5
GNWVTC
genome polyprotein - hepatitis C virus
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
C;Accession: A38465
R;Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; F
J. Virol. 65, 1105-1113, 1991
A;Title: Structure and organization of the hepatitis C virus genome isolated from human c
A;Reference number: A38465; MUID:91140698; PMID:1847440
A;Accession: A38465
A;Molecule type: genomic RNA
A;Residues: 1-3010 <TAK>
A;Cross-references: EMBL:M58335; NID:g329770; PIDN:AAA72945.1; PID:g329771
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,224

  Query Match           93.0%;  Score 1478;  DB 1;  Length 3010;
  Best Local Similarity   92.1%;  Pred. No. 3.5e-119;
  Matches  279;  Conservative   11;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||:|||||||||| ||||||| |||||||||||| ||||||:|:|||||:||  ||
Db        904 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWPRAG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |||||||||||||| |||||||||:||||||: ||:|
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGRGL 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              |||||||| |||||||||||||||  ||||||:|||||||||||||||||||||||||||
Db       1084 GAGSKTLAAPKGPITQMYTNVDQDLVGWPKPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPFGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 6
S40770
genome polyprotein - hepatitis C virus
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C;Accession: S40770; PC1285
R;Okamoto, H.
submitted to the EMBL Data Library, March 1992
A;Reference number: S40770
A;Accession: S40770
A;Molecule type: genomic RNA
```

A;Residues: 1-3011 <OKA>
A;Cross-references: EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g221587
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,
Jpn. J. Exp. Med. 60, 167-177, 1990
A;Title: The 5'-terminal sequence of the hepatitis C virus genome.
A;Reference number: PC1284; MUID:91013116; PMID:2170712
A;Accession: PC1285
A;Molecule type: genomic RNA
A;Residues: 1-513 <OK2>
A;Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512
A;Experimental source: isolate HC-J1
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 88.3%; Score 1403; DB 1; Length 3011;
Best Local Similarity 84.8%; Pred. No. 1.1e-112;
Matches 257; Conservative 23; Mismatches 23; Indels 0; Gaps 0;

```
Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              | : ||||||| |||:| | | ||  |||||||| :|| ||||||||:|||||:|||| |
Db        904 ASLLKVPYFVRVQGLLRICALARKMVGGHYVQMAIIKLGALTGTYVYNHLTPLRDWAHNG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:|| || |:|||||||||||||:|||||||:||||||||||     :||
Db        964 LRDLAVAVEPVVFSQMETKLITWGADTAACGDIINGLPVSARKGREILLGPADGMVSKGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              |||||||||:|||||||||||||||||||||||||||||:|||| |:|||||:|||||||:|
Db       1024 RLLAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              |||::|:| |||||: |||||||||||||| || ||||:|||||||||||||||||||||||
Db       1084 GAGTRTIASPKGPVIQMYTNVDQDLVGWPAPQGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              |||||||||||||:|||||||||||||:|| ||||||||||||||||||||||||||:||
Db       1144 RRGDSRGSLLSPRPISYLKGSSGGPLLCPAGHVVGIFRAAVCTRGVAKAVDFIPVESLET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206
```

RESULT 7
GNWVC3
genome polyprotein - hepatitis C virus (strain HCV-1)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001
C;Accession: A39166; PQ0403; PQ0404
R;Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
A;Title: Genetic organization and diversity of the hepatitis C virus.
A;Reference number: A39166; MUID:91172826; PMID:1848704
A;Accession: A39166
A;Molecule type: mRNA
A;Residues: 1-3011 <CHO>
A;Cross-references: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874
R;Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L
J. Gen. Virol. 73, 1131-1141, 1992

A;Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to ex
A;Reference number: PQ0393; MUID:92268871; PMID:1316939
A;Accession: PQ0403
A;Molecule type: genomic RNA
A;Residues: 1577-1633 <CHA>
A;Cross-references: DDBJ:D10128
A;Experimental source: isolates E-b16
A;Accession: PQ0404
A;Status: preliminary
A;Molecule type: genomic RNA
A;Residues: 1577-1633 <CH2>
A;Experimental source: isolates E-b17
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,224

Query Match 88.0%; Score 1398; DB 1; Length 3011;
Best Local Similarity 84.5%; Pred. No. 2.9e-112;
Matches 256; Conservative 24; Mismatches 23; Indels 0; Gaps 0;

```
Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              | : ||||||| |||:| | | ||  |||||||  :|| ||||||||:|||||:|||| |
Db        904 ASLLKVPYFVRVQGLLRFCALARKMIGGHYVQMVIIKLGALTGTYVYNHLTPLRDWAHNG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:|| || |:|||||||||||||:|||||||||||||||||     :||
Db        964 LRDLAVAVEPVVFSQMETKLITWGADTAACGDIINGLPVSARRGREILLGPADGMVSKGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              |||||||||:|||||||||||||||||||||||||||||:|||| |:|||||:|||||||:|
Db       1024 RLLAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              |||::|:| |||||: |||||||||||||| || |:||:|||||||||||||||||||||||
Db       1084 GAGTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              |||||||||||||:|||||||||||||:|||||||||||||||||||||||||||||::||
Db       1144 RRGDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206
```

RESULT 8
GNWVCH
genome polyprotein - hepatitis C virus (strain H)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructur
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: A36814; A41546
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A;Description: Genomic structure of the human prototype strain H of hepatitis C virus: cc
A;Reference number: A36814
A;Accession: A36814
A;Molecule type: genomic RNA
A;Residues: 1-3011 <INC>

```
A;Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A;Title: Genomic structure of the human prototype strain H of hepatitis C virus: compari
A;Reference number: A41546; MUID:92052256; PMID:1658800
A;Contents: annotation
A;Note: neither amino acid nor nucleotide sequence is given
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

  Query Match             87.2%;  Score 1385;  DB 1;  Length 3011;
  Best Local Similarity   83.8%;  Pred. No. 3.9e-111;
  Matches  254;  Conservative   26;  Mismatches   23;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              | : ||||||| |||:| | | || ||||||||| :|| ||||| ||:|| ||:|||| |
Db        904 ASLLKVPYFVRVQGLLRICALARKIAGGHYVQMAIIKLGALTGTCVYNHLAPLRDWAHNG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:|| || |:|||||||||||:|||||||:|||||||     :||
Db        964 LRDLAVAVEPVVFSRMETKLITWGADTAACGDIINGLPVSARRGQEILLGPADGMVSKGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              |||||||||:|||||||||||||||||||||||||:|||||:|||||:|||||||:|
Db       1024 RLLAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTATQTFLATCINGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              |||::|:| |||| | |||||||||||| || |:||:|||||||||||||||||||||||
Db       1084 GAGTRTIASPKGPVIQTYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              |||||||||||||:|||||||||||||:|||||:|||||||||||||||||||::||
Db       1144 RRGDSRGSLLSPRPISYLKGSSGGPLLCPTGHAVGLFRAAVCTRGVAKAVDFIPVENLET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 9
JC5620
genome polyprotein - hepatitis C virus (isolate EUH1480)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C;Accession: JC5620
R;Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
Biochem. Biophys. Res. Commun. 236, 44-49, 1997
A;Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant
A;Reference number: JC5620; MUID:97366593; PMID:9223423
A;Accession: JC5620
A;Molecule type: mRNA
A;Residues: 1-3014 <CHA>
A;Cross-references: GB:Y13184
A;Experimental source: genotype 5a, which predominates in South Africa
A;Note: the translation of the nucleotide sequence is not complete in this paper
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;384-408/Region: hypervariable #status predicted
F;390-730/Product: nonstructural protein NS1 #status predicted <NS1>
F;731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
F;1008-1616/Product: hepacivirin #status predicted <NS3>
F;1231-1238/Region: nucleotide-binding motif A (P-loop)
F;1313-1318/Region: nucleotide-binding motif B
F;1317-1320/Region: DEXH motif
F;1617-1863/Product: nonstructural protein NS4a #status predicted <N4A>
F;1864-2014/Product: nonstructural protein NS4b #status predicted <N4B>
F;2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
F;2210-2249/Region: interferon sensitivity determining #status predicted

  Query Match             77.7%;  Score 1234;  DB 1;  Length 3014;
  Best Local Similarity   72.8%;  Pred. No. 4.7e-98;
  Matches  219;  Conservative   43;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          3 ITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAGLR 62
              :||||||:||: |:| |:| :    | ||| | : |  |||||:|||| |::||| :|||
Db        907 LTKVPYFLRARALLRLCLLAKHLVYGKYVQAALLHLGRLTGTYIYDHLAPMKDWAASGLR 966

Qy         63 DLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGWRL 122
              :| || ||::|| || |:||||||||||:|:: |||||||||| |||||: :  ||||
Db        967 ELTVATEPIVFSAMETKVITWGADTAACGNILAVLPVSARRGREIFLGPADDIKTSGWRL 1026

Qy        123 LAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFHGA 182
              |||||||:|||||:|| |: ||||||||: ||||| :|||||:||  |:||| ||:||||
Db       1027 LAPITAYAQQTRGVLGAIVLSLTGRDKNEAEGEVQFLSTATQTFLGICINGVMWTLFHGA 1086

Qy        183 GSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVRRR 242
              ||||||||||: |||||||:||||| :||| |:| |||||:|||||||||||||| |||
Db       1087 GSKTLAGPKGPVVQMYTNVDKDLVGWPSPPGKGSLTRCTCGSADLYLVTRHADVIPARRR 1146

Qy        243 GDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTM 302
              ||:| |||||||:||||||||::||||| ||:|||||||||||::|:|||::||||
Db       1147 GDTRASLLSPRPISYLKGSSGGPIMCPSGHVVGVFRAAVCTRGVAKALEFVPVENLETTM 1206

Qy        303 R 303
              |
Db       1207 R 1207


RESULT 10
JQ1303
genome polyprotein - hepatitis C virus (isolate HC-J6)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
C;Accession: JQ1303
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.;
J. Gen. Virol. 72, 2697-2704, 1991
A;Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a huma
A;Reference number: JQ1303; MUID:92044440; PMID:1658196
A;Accession: JQ1303
A;Molecule type: genomic RNA
A;Residues: 1-3033 <OKA>
A;Cross-references: GB:D00944; NID:g221650; PIDN:BAA00792.1; PID:g221651
A;Experimental source: isolate HC-J6 from a Japanese individual
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transme
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F;1011-1619/Product: hepacivirin #status predicted <NS3>
F;1316-1321/Region: nucleotide-binding motif B
F;1320-1323/Region: DEXH motif
F;1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>
F;1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>
F;2018-3033/Product: nonstructural protein NS5 #status predicted <NO5>
```

F;196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28

Query Match 73.8%; Score 1172; DB 1; Length 3033;
Best Local Similarity 69.4%; Pred. No. 1.1e-92;
Matches 209; Conservative 44; Mismatches 48; Indels 0; Gaps 0;

```
Qy          3 ITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAGLR 62
              :|:|||||||  |:| | :||  ||| ||||  : |   ||||:||||||: |||  |||
Db        910 LTRVPYFVRAHALLRMCTMVRHLAGGRYVQMVLLALGRWTGTYIYDHLTPMSDWAANGLR 969

Qy         63 DLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGWRL 122
              ||||||||:||| || |:| |||:|||||||: ||||||| |||:|||||| :  :|| |
Db        970 DLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPADGYTSKGWSL 1029

Qy        123 LAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFHGA 182
              |||||||:|||||||| |: |:||||| :  ||:||:|| ||||| | ::|| |||:|||
Db       1030 LAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTTISGVLWTVYHGA 1089

Qy        183 GSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVRRR 242
              |:||||| :||:||||:: : ||||| :||| :|: ||||: |||||||:||||| |||
Db       1090 GNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRNADVIPARRR 1149

Qy        243 GDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTM 302
              || ||:|||||:| |||||||||:||| |||||:||||||:|||||::||||||:::
Db       1150 GDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFIPVETLDIVT 1209

Qy        303 R 303
              |
Db       1210 R 1210
```

RESULT 11
GNWVJ8
genome polyprotein - hepatitis C virus (strain HC-J8)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: A40250; PQ0397; PQ0559
R;Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.;
Virology 188, 331-341, 1992
A;Title: Full-length sequence of a hepatitis C virus genome having poor homology to repc
A;Reference number: A40250; MUID:92230232; PMID:1314459
A;Accession: A40250
A;Molecule type: genomic RNA
A;Residues: 1-3033 <OKA>
A;Cross-references: GB:D10988; GB:D01221; NID:g221608; PIDN:BAA01761.1; PID:g221609
R;Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L
J. Gen. Virol. 73, 1131-1141, 1992
A;Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
A;Reference number: PQ0393; MUID:92268871; PMID:1316939
A;Accession: PQ0397
A;Molecule type: genomic RNA
A;Residues: 2678-2754 <CHA>
A;Cross-references: DDBJ:D10134
A;Experimental source: isolate E-b12
R;Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohnc
Biochem. Biophys. Res. Commun. 181, 279-285, 1991
A;Title: Distribution of plural HCV types in Japan.
A;Reference number: PQ0554; MUID:92068204; PMID:1720309
A;Accession: PQ0559
A;Molecule type: mRNA
A;Residues: 2678-2729 <KAT>
A;Cross-references: GB:D10562; GB:D90518; NID:g221523; PIDN:BAA01418.1; PID:g221524
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F;1011-1619/Product: hepacivirin #status predicted <NS3>
F;1234-1241/Region: nucleotide-binding motif A (P-loop)
F;1316-1321/Region: nucleotide-binding motif B
F;1320-1323/Region: DEXH motif
F;1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>
F;1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>
F;2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,235

Query Match 72.4%; Score 1150; DB 1; Length 3033;
Best Local Similarity 67.3%; Pred. No. 9e-91;
Matches 204; Conservative 47; Mismatches 52; Indels 0; Gaps 0;

```
Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              | : ::||||||  |:| | ||:  ||  |:||  : :   ||||:||||:||  ||  |
Db        908 ASLLRIPYFVRAHALLRVCTLVKHLAGARYIQMLLITIGRWTGTYIYDHLSPLSTWAAQG 967

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||:|||||:|| || |:| |||:| |||||: |||||| |||:|||||| :  :||
Db        968 LRDLAIAVEPVVFSPMEKKVIVWGAETVACGDILHGLPVSARLGREVLLGPADGYTSKGW 1027

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              :|||||||:|||||||| |: ||||||||:  |:|||:|: ||:|| | ::|| |||:|
Db       1028 KLLAPITAYTQQTRGLLGAIVVSLTGRDKNEQAGQVQVLSSVTQTFLGTSISGVLWTVYH 1087

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              |||:|||||||||:|||||: : ||||| :||| :|: |||||: |||||||:|||||||
Db       1088 GAGNKTLAGPKGPVTQMYTSAEGDLVGWPSPPGTKSLDPCTCGAVDLYLVTRNADVIPVR 1147

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              |: | ||:|||||:| ||||||||:||  |||||:|||||| |||||::||||||::
Db       1148 RKDDRRGALLSPRPLSTLKGSSGGPVLCSRGHAVGLFRAAVCARGVAKSIDFIPVESLDV 1207

Qy        301 TMR 303
              |||
Db       1208 ATR 1210
```

RESULT 12
T08841
polyprotein - douroucouli hepatitis GB virus A
C;Species: douroucouli hepatitis GB virus A
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C;Accession: T08841
R;Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
J. Gen. Virol. 79, 41-45, 1998
A;Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
A;Reference number: Z16486; MUID:98120818; PMID:9460920
A;Accession: T08841
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-3005 <ERK>
A;Cross-references: EMBL:AF023425; NID:g2828599; PIDN:AAC40502.1; PID:g2828600
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: polyprotein

Query Match 25.0%; Score 397.5; DB 2; Length 3005;
Best Local Similarity 34.8%; Pred. No. 1.3e-25;
Matches 95; Conservative 46; Mismatches 119; Indels 13; Gaps 5;

```
Qy         34 AFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDI 93
              ||::     |  :: |   :   | | | || ||:|||  :  :  |:   | | |||
Db        887 AFVRRLERRGVTLFQHCGQVSXGAAAILXDLGVALEPVSVTARDCYIVRDAARTLACGQR 946

Qy         94 ISGLPVSARRGREILLG--PADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQ 151
              : |||| |||| |:|:|  |:      |:   ||:    |: |    : ||: |||: :
Db        947 VEGLPVVARRGDEVLVGVFPSVRALPPGFVPTAPVVV-MQRGLGFFSVVKTSMLGRDERE 1005

Qy        152 VEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAP 211
              || : |: |:|   : |||||| :| |||: ::||||| ||:   : :   |:  : |
Db       1006 HEGSIVVLGTSTTRSMGTCVNGVMYTTFHGSNARTLAGPVGPVNCRWWSPSDDVAVYPLP 1065

Qy        212 PGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSG 271
              ||  : || ||:  :: :   |     |     |  |  :|  :|||| |:||  |
Db       1066 SGASCLEPCKCGTQSVWCIRN--DGALCHGRLSKLVELDLPTEISDFRGSSGSPILCDEG 1123
```

```
Qy        272 HAVGIFRAAVCTRGV-------AKAVDFIPVES 297
              | ||:   :|  |||       |  : :| :|
Db       1124 HVVGMM-VSVLHRGVKVTGVRYVKPWETLPKDS 1155
```

RESULT 13
T08839
polyprotein - marmoset hepatitis GB virus A
C;Species: marmoset hepatitis GB virus A
C;Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 17-Nov-2000
C;Accession: T08839
R;Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
J. Gen. Virol. 79, 41-45, 1998
A;Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
A;Reference number: Z16486; MUID:98120818; PMID:9460920
A;Accession: T08839
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: genomic RNA
A;Residues: 1-2970 <ERK>
A;Cross-references: EMBL:AF023424; NID:g2828597; PIDN:AAC40501.1; PID:g2828598
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: polyprotein

  Query Match 21.5%; Score 341; DB 2; Length 2970;
  Best Local Similarity 30.5%; Pred. No. 1e-20;
  Matches 85; Conservative 46; Mismatches 96; Indels 52; Gaps 9;

```
Qy         56 WAHAG---------LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGRE 106
              :||||          ||    |:|||      :  ::   | | :||  : | || |||| |
Db        891 YAHAGQVTRRTAEQLRQWGFALEPVAVHPEDCAMVRDAARTLSCGQSVHGKPVVARRGDE 950

Qy        107 ILLGPADNFEGQGWRL------LAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVS 160
              :|:|  :      | |       ||:  :    :|  | : ||:|| |: :  | | |:
Db        951 VLIGVLNGV----WELPPGFVPTAPVVVH-HHGKGFFGVVKTSMTGWDETEHVGNVVVLG 1005

Qy        161 TATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPC 220
              |:|   : |||||| :| :||: ::|||   ||:   : :   |:  :  | ||: : ||
Db       1006 TSTTRSMGTCVNGVMYTTYHGSNARTLAAQMGPVNSRWWSASDDVAVYPLPVGAKCLEPC 1065

Qy        221 TCGSSDLYLVTRHADVIPVRRRGDSRGSLLS-----------PRPVSYLKGSSGGPLLCP 269
              |    ::::          | |  |:|              |  :   :|||| |:||
Db       1066 KCQPQGVWVI-----------RND--GALCHGTLGRTVELDLPAELCDFRGSSGSPILCD 1112

Qy        270 SGHAVGIFRAAVCTRG-------VAKAVDFIPVESMETT 301
              |||||:   :|  ||        |  : :| |::  |
Db       1113 EGHAVGML-ISVLHRGSRVTGIRYTKPWETLPREAITHT 1150
```

RESULT 14
B46642
DNA-directed DNA polymerase (EC 2.7.7.7) alpha/DNA primase (EC 2.7.7.-) complex 68K chai
C;Species: Mus musculus (house mouse)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Jun-2000
C;Accession: B46642
R;Miyazawa, H.; Izumi, M.; Tada, S.; Takada, R.; Masutani, M.; Ui, M.; Hanaoka, F.
J. Biol. Chem. 268, 8111-8122, 1993
A;Title: Molecular cloning of the cDNAs for the four subunits of mouse DNA polymerase al
A;Reference number: A46642; MUID:93216788; PMID:8463324
A;Accession: B46642
A;Status: preliminary
A;Molecule type: mRNA; protein
A;Residues: 1-600 <MIY>
A;Cross-references: GB:D13546; NID:g303658; PIDN:BAA02746.1; PID:g303659
A;Experimental source: FM3A cells
A;Note: sequence extracted from NCBI backbone (NCBIN:129148, NCBIP:129149)
C;Keywords: nucleotidyltransferase

  Query Match 6.4%; Score 101; DB 2; Length 600;
  Best Local Similarity 24.8%; Pred. No. 0.85;
  Matches 55; Conservative 34; Mismatches 71; Indels 62; Gaps 12;

```
Qy         15 LIRACMLVRKAAGGHYVQM-AFMKLAALT-------GTYVYDHL---------TPLQDWA 57
              |   |:| |:   |   :: ||   | |        :: |: |         :  :|
Db         27 LAELCVLYRQTEDGMVSELIAFCTSAGKTCLTVDILNSFEYEVLNKKLSKAWHSASKDSG 86

Qy         58 HAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDI--ISGLP--------VSARRGREI 107
              ||| ||: |::: :| :: | : :     | : | :  :|  |         |:||  |:
Db         87 HAGTRDI-VSIQELIEAEEEEETLLSSYTTPSKGPLKRVSSTPETPLTKRSVAARSPRQ- 144

Qy        108 LLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFL 167
              || |:           :|    ||:         || | |       |||     : |
Db        145 LLSPSS---------FSPSATPSQK--------YTSRTNR------GEVVTTFGSAQ--- 178

Qy        168 ATCVNGVCWTVFHGAGSKTL--AGPKGPITQMYTNVDQDLVG 207
                |: |:   |:|| :|   |   |:|  | : | |:|
Db        179 -----GLSWSGRGGSGSVSLKVVGDPEPLTGSYKAMFQQLMG 215
```

RESULT 15
G87392
conserved hypothetical protein CC1155 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: G87392
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolona
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: G87392
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-353 <STO>
A;Cross-references: GB:AE005673; NID:g13422473; PIDN:AAK23139.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1155

  Query Match 6.3%; Score 99.5; DB 2; Length 353;
  Best Local Similarity 21.8%; Pred. No. 0.59;
  Matches 72; Conservative 36; Mismatches 114; Indels 109; Gaps 13;

```
Qy          7 PYFVRAQGLIRACMLVRKAA---------GGHYV-------QMAFMKLAALTGTYVYDHL 50
              | | | ||      | |::|          |: :        :| | | || |  :
Db         65 PIAVLAGGLFAFSQLARESAIVAMRASGISGYRIVGMAVPAAVAVMLLDALCGQVLAPRA 124

Qy         51 TP-LQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILL 109
               | | ||          :|   |||  :      :  ||: :| |  |:|: :  |  : ||:
Db        125 DPTLADW-WRNTTPVAERKEPVPRTFRAGADLVIGANASADGRTITGVTIFRRDSKGILV 183

Qy        110 ----GPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQS 165
                  ||  ::|: | |  | |                          :  |::   |||
Db        184 EKVEAPAARYDGKAWTLEQPKT----------------------TRFAGDLSQASTAA-- 219

Qy        166 FLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSS 225
                          |                    | :     | | |  : | | |
Db        220 -------ATSWP----------------TALRPQDVQGLFGDDSMPTAAS--------- 246

Qy        226 DLYLVTRHADVIPVRRRGDSRGSLLSPRPVSY----LKGSSGGP------LLCPSGHAVG 275
                            ||  :: |   | || |:    |: :   |       ||  :  |:
Db        247 ------------ARRALENGG---SDRPESFYATHLQAAFASPFVSLVMLLLSAPVALA 290

Qy        276 IFR----AAVCTRGVAKAVDFIPVESMETTM 302
              ||    | : | |:|  : |:     | | :
Db        291 NFRSGQGAVLLTGGLAAGLMFLVANGMLTAL 321
```

Search completed: May 6, 2004, 09:37:15
Job time : 10.8681 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:09:55 ; Search time 6.32568 Seconds
(without alignments)
2494.160 Million cell updates/sec

Title: US-10-650-585-10
Perfect score: 1589
Sequence: 1 AGITKVPYFVRAQGLIRACM.....RGVAKAVDFIPVESMETTMR 303

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1532	96.4	3010	1	POLG_HCVJT	Q00269 h genome po
2	1527	96.1	3010	1	POLG_HCVJA	P26662 h genome po
3	1514	95.3	3010	1	POLG_HCVTW	P29846 h genome po
4	1478	93.0	3010	1	POLG_HCVBK	P26663 h genome po
5	1398	88.0	3011	1	POLG_HCV1	P26664 h genome po
6	1385	87.2	3011	1	POLG_HCVH	P27958 h genome po
7	1172	73.8	3033	1	POLG_HCVJ6	P26660 h genome po
8	1150	72.4	3033	1	POLG_HCVJ8	P26661 h genome po
9	101	6.4	600	1	DPO2_MOUSE	P33611 mus musculu
10	93	5.9	660	1	VST2_HEVBU	P29326 hepatitis e
11	93	5.9	660	1	VST2_HEVPA	P33426 hepatitis e
12	92.5	5.8	706	1	TRFE_HORSE	P27425 equus cabal
13	89.5	5.6	659	1	VST2_HEVME	Q03500 hepatitis e
14	88	5.5	3414	1	POLG_TBEVW	P14336 t genome po
15	87	5.5	3412	1	POLG_TBEVS	P07720 t genome po
16	87	5.5	3414	1	POLG_TBEVH	Q01299 t genome po
17	86.5	5.4	470	1	NRAM_IAWHM	P05803 influenza a
18	86	5.4	434	1	TOLB_CHLTE	Q8keq0 chlorobium
19	85	5.3	470	1	NRAM_IATRA	P03472 influenza a
20	85	5.3	730	1	HELS_METMA	Q8pzr7 methanosarc
21	84.5	5.3	347	1	MDHM_EUCGU	P46487 eucalyptus
22	84.5	5.3	1705	1	PTPV_MOUSE	P70289 mus musculu
23	84	5.3	309	1	UCP2_RAT	P56500 rattus norv
24	84	5.3	339	1	GPDA_COREF	Q8fpr0 corynebacte
25	84	5.3	470	1	NRAM_IARUE	P31510 influenza a
26	83.5	5.3	538	1	DAC_ACTSP	P39045 actinomadur
27	83.5	5.3	854	1	PWP2_SCHPO	Q9c1x1 schizosacch
28	83	5.2	485	1	VST2_HEVRH	Q00270 hepatitis e
29	83	5.2	660	1	VST2_HEVMY	Q04611 hepatitis e
30	82.5	5.2	453	1	NRAM_IAWIL	P03470 influenza a
31	82	5.2	309	1	UCP2_MOUSE	P70406 mus musculu
32	82	5.2	403	1	PGK_CHLMU	Q9pln4 chlamydia m
33	82	5.2	612	1	AMYG_ASPOR	P36914 aspergillus
34	81.5	5.1	398	1	TRMU_AGRT5	Q8u9m5 agrobacteri
35	81.5	5.1	1022	1	CA26_CHICK	P15988 gallus gall
36	81	5.1	209	1	PAAD_PSEAE	Q9hx08 pseudomonas
37	81	5.1	350	1	PE24_ARATH	Q9zv04 arabidopsis
38	81	5.1	470	1	NRAM_IAUSS	P03469 influenza a
39	81	5.1	730	1	HELS_METAC	Q8tl39 methanosarc
40	80	5.0	309	1	UCP2_HUMAN	P55851 homo sapien
41	80	5.0	326	1	PANE_RHILO	Q987n5 rhizobium l
42	80	5.0	961	1	ATCU_YERPE	Q8zca7 yersinia pe
43	80	5.0	3491	1	ERY1_SACER	Q03131 saccharopol
44	79.5	5.0	339	1	CRTB_RHOCA	P17056 rhodobacter
45	79.5	5.0	453	1	GAG_AVIMD	P06444 avian myelo

ALIGNMENTS

RESULT 1
POLG_HCVJT
ID POLG_HCVJT STANDARD; PRT; 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals.";
RL Virus Res. 23:39-53(1992).
CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function. NS3
CC and NS5 may play a role in the viral RNA replication.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; D11168; BAA01943.1; -.
DR PIR; A45573; A45573.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.

```
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT   CHAIN       730   1006       NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3010       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE    1083   1083      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1107   1107      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1165   1165      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND     1230   1237      ATP (POTENTIAL).
FT   SITE        1316   1319      DECH BOX.
FT   CARBOHYD     196    196      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     209    209      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     234    234      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     250    250      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     305    305      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     417    417      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     423    423      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     430    430      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     448    448      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     532    532      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     540    540      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     556    556      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     576    576      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     623    623      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     645    645      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2041   2041      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2077   2077      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2240   2240      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2529   2529      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2788   2788      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3010 AA;  326573 MW;  94A1C77435D642BB CRC64;

  Query Match             96.4%;  Score 1532;  DB 1;  Length 3010;
  Best Local Similarity   95.4%;  Pred. No. 1.6e-124;
  Matches  289;  Conservative    6;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              | || :|||||||||||||||||| ||||||||||||||||||||||||||||||||||||
```

```
Db        904 AAITAMPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:||||| |||||||||||||||| |||||||||||||||||: |||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGPADSIEGQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              |||||||||:|||||||||:||||||||||||||||||||||||||||||||||||||||
Db       1024 RLLAPITAYAQQTRGLLGCIVTSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              |||||||||||||||||||||||||||| |||||||:||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWHAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 RRGDGRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206
```

```
RESULT 2
POLG_HCVJA
ID   POLG_HCVJA     STANDARD;      PRT;  3010 AA.
AC   P26662;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate Japanese) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91088550; PubMed=2175903;
RA   Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA   Sugimura T., Shimotohno K.;
RT   "Molecular cloning of the human hepatitis C virus genome from
RT   Japanese patients with non-A, non-B hepatitis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RN   [2]
RP   DISCUSSION OF SEQUENCE.
RX   MEDLINE=91192160; PubMed=1849488;
RA   Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraiso K.,
RA   Ohkoshi S., Shimotohno K.;
RT   "Molecular structure of the Japanese hepatitis C viral genome.";
RL   FEBS Lett. 280:325-328(1991).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D90208; BAA14233.1; -.
DR   PIR; A39253; GNWVCJ.
DR   HSSP; P26663; 1JXP.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT   CHAIN       730   1006       NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3010       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
FT   SITE       1316   1319       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2041   2041       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2077   2077       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2240   2240       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2788   2788       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3010 AA;  327017 MW;  AA993794F46DB185 CRC64;

  Query Match             96.1%;  Score 1527;  DB 1;  Length 3010;
  Best Local Similarity   94.4%;  Pred. No. 4.4e-124;
  Matches  286;  Conservative   11;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||:|||||||||||||||||| |||||||||||||||||||||||||||:||||||
Db        904 AGITRVPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLRDWAHAG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |:||||||||||||||||||||:||||||||:|  |||
Db        964 LRDLAVAVEPVVFSDMETKLITWGADTAACGDIISGLPVSARRGKEILLGPADSFGEQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||:|||||:||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVDGEVQVLSTATQSFLATCVNGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWPAPPGARSMTPCTCGSSDLYLVTRHADVVPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              |||||||||||||:|||||||||||||||||| ||||||||||||||||||||||||||
Db       1144 RRGDSRGSLLSPRPISYLKGSSGGPLLCPSGHVVGIFRAAVCTRGVAKAVDFIPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 3
POLG_HCVTW
ID   POLG_HCVTW     STANDARD;      PRT;  3010 AA.
AC   P29846;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate Taiwan) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=31645;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92230206; PubMed=1314449;
RA   Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
RT   "The Taiwanese hepatitis C virus genome: sequence determination and
RT   mapping the 5' termini of viral genomic and antigenomic RNA.";
RL   Virology 188:102-113(1992).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M84754; -; NOT_ANNOTATED_CDS.
DR   PIR; A40244; GNWVTW.
DR   PDB; 1N64; 25-FEB-03.
DR   PDB; 1NS3; 08-APR-98.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CORE PROTEIN (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT   CHAIN       730   1006       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3010       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE    1083   1083      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1107   1107      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1165   1165      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND     1230   1237      ATP (POTENTIAL).
FT   SITE        1316   1319      DECH BOX.
FT   CARBOHYD     196    196      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     209    209      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     233    233      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     234    234      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     250    250      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     305    305      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     417    417      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     423    423      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     430    430      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     448    448      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     532    532      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     540    540      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     556    556      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     576    576      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     623    623      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     645    645      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2041   2041      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2077   2077      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2240   2240      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2529   2529      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2788   2788      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3010 AA;  327047 MW;  AAD267D55CDFE215 CRC64;

  Query Match             95.3%;  Score 1514;  DB 1;  Length 3010;
  Best Local Similarity   93.1%;  Pred. No. 5.9e-123;
  Matches  282;  Conservative   12;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||:|::|||||||||||||||||||||||||||| |||||||||||||||||||||| |
Db        904 AGLTRIPYFVRAQGLIRACMLVRKAAGGHYVQMALMKLAALTGTYVYDHLTPLQDWAHTG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:|||| |||||||||||||| |||||||||||||||||||: ||:||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGPADSLEGRGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              |||||||||:|||||| |||||||||||||||||||||||||||||||:|||||||:|
Db       1024 RLLAPITAYAQQTRGLFGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCINGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              |||||||||||||||||||||||||||| || ||||:|||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWHAPQGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              |||||||||||||||:|||||||||||||| |||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPISYLKGSSGGPLLCPSGHVVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 4
POLG_HCVBK
ID   POLG_HCVBK     STANDARD;      PRT;  3010 AA.
AC   P26663;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate BK) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11105;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91140698; PubMed=1847440;
RA   Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RA   Onishi E., Andoh T., Yoshida I., Okayama H.;
RT   "Structure and organization of the hepatitis C virus genome isolated
RT   from human carriers.";
RL   J. Virol. 65:1105-1113(1991).
RN   [2]
RP   SEQUENCE OF 1487-1500.
RX   MEDLINE=96235224; PubMed=8647104;
RA   Borowski P., Heiland M., Oehlmann K., Becker B., Korneteky L.;
RT   "Non-structural protein 3 of hepatitis C virus inhibits
```

```
RT   phosphorylation mediated by cAMP-dependent protein kinase.";
RL   Eur. J. Biochem. 237:611-618(1996).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX   MEDLINE=97015088; PubMed=8861916;
RA   Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
RA   Moomaw E.W., Adachi T., Hostomska Z.;
RT   "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RT   trypsin-like fold and a structural zinc binding site.";
RL   Cell 87:331-342(1996).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX   MEDLINE=98227846; PubMed=9568891;
RA   Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
RA   Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
RT   "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
RT   virus: a 2.2-A resolution structure in a hexagonal crystal form.";
RL   Protein Sci. 7:837-847(1998).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M58335; AAA72945.1; -.
DR   PIR; A38465; GNWVTC.
DR   PDB; 1A1Q; 25-MAR-98.
DR   PDB; 1JXP; 14-JAN-98.
DR   PDB; 1NS3; 08-APR-98.
DR   PDB; 1C2P; 15-NOV-00.
DR   PDB; 1CSJ; 08-NOV-99.
DR   PDB; 1GX5; 09-APR-02.
DR   PDB; 1GX6; 10-APR-02.
DR   PDB; 1QUV; 26-JUN-00.
DR   PDB; 8OHM; 20-APR-99.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT   CHAIN       730   1006       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3010       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM.
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM.
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM.
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
FT   SITE       1316   1319       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2041   2041       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2077   2077       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2240   2240       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2529   2529       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2788   2788       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   STRAND     1031   1035
FT   HELIX      1039   1047
FT   STRAND     1050   1050
FT   STRAND     1059   1063
FT   STRAND     1068   1074
FT   TURN       1075   1076
FT   STRAND     1077   1081
FT   HELIX      1082   1085
FT   TURN       1086   1087
FT   STRAND     1090   1092
FT   TURN       1093   1094
FT   STRAND     1095   1097
FT   STRAND     1101   1103
FT   TURN       1104   1107
FT   STRAND     1108   1112
FT   STRAND     1120   1120
FT   STRAND     1122   1122
FT   STRAND     1129   1133
FT   TURN       1135   1136
FT   STRAND     1139   1144
FT   STRAND     1149   1157
FT   HELIX      1158   1161
FT   TURN       1162   1163
FT   TURN       1165   1166
FT   STRAND     1168   1171
FT   TURN       1172   1174
```

```
FT   STRAND      1175   1186
FT   TURN        1187   1188
FT   STRAND      1189   1197
FT   HELIX       1198   1202
FT   TURN        1203   1204
FT   STRAND      1680   1688
SQ   SEQUENCE   3010 AA;  327189 MW;  F8422D5ECCFDFD9C CRC64;

  Query Match             93.0%;  Score 1478;  DB 1;  Length 3010;
  Best Local Similarity   92.1%;  Pred. No. 7.8e-120;
  Matches  279;  Conservative   11;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||:|||||||||||| ||||||| ||||||||||||| ||||||:|:|||||:||  ||
Db        904 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWPRAG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:|||||  |||||||||||||||| |||||||||:||||||||: ||:|
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGRGL 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              |||||||| |||||||||||||||||   ||||||:|||||||||||||||||||||||
Db       1084 GAGSKTLAAPKGPITQMYTNVDQDLVGWPKPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||| ||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPFGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 5
POLG_HCV1
ID   POLG_HCV1      STANDARD;      PRT;  3011 AA.
AC   P26664;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate 1) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11104;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91172826; PubMed=1848704;
RA   Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA   Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
RA   Bradley D.W., Kuo G., Houghton M.;
RT   "Genetic organization and diversity of the hepatitis C virus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M62321; AAA45676.1; -.
DR   PIR; A39166; GNWVC3.
DR   PDB; 1A1V; 16-FEB-99.
DR   PDB; 1HEI; 25-NOV-98.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT   CHAIN       730   1006       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3011       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
FT   SITE       1316   1319       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    476    476       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2041   2041       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2077   2077       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2240   2240       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2364   2364       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2789   2789       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3011 AA;  327197 MW;  65F8C9447FCE5AF9 CRC64;

  Query Match             88.0%;  Score 1398;  DB 1;  Length 3011;
  Best Local Similarity   84.5%;  Pred. No. 7e-113;
  Matches  256;  Conservative   24;  Mismatches   23;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              | : ||||||| |||:|  | | ||  |||||||  :|| ||||||||:|||||:|||| |
Db        904 ASLLKVPYFVRVQGLLRFCALARKMIGGHYVQMVIIKLGALTGTYVYNHLTPLRDWAHNG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:|| || |:|||||||||||||:|||||||||||||||||    :||
Db        964 LRDLAVAVEPVVFSQMETKLITWGADTAACGDIINGLPVSARRGREILLGPADGMVSKGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              |||||||||:|||||||||||||||||||||||||:|||| |:|||||:|||||||:|
Db       1024 RLLAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              |||::|:| ||||: |||||||||||| || |:||:|||||||||||||||||||||||
Db       1084 GAGTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              |||||||||||||:|||||||||||||:||||||||||||||||||||::||
Db       1144 RRGDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 6
POLG_HCVH
ID   POLG_HCVH      STANDARD;      PRT;  3011 AA.
AC   P27958;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate H) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11108;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92052256; PubMed=1658800;
RA   Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA   Prince A.M.;
RT   "Genomic structure of the human prototype strain H of hepatitis C
RT   virus: comparison with American and Japanese isolates.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX   MEDLINE=97331322; PubMed=9187654;
RA   Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT   "Structure of the hepatitis C virus RNA helicase domain.";
RL   Nat. Struct. Biol. 4:463-467(1997).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX   MEDLINE=98154321; PubMed=9493270;
RA   Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA   Murcko M.A., Lin C., Caron P.R.;
RT   "Hepatitis C virus NS3 RNA helicase domain with a bound
RT   oligonucleotide: the crystal structure provides insights into the mode
RT   of unwinding.";
RL   Structure 6:89-100(1998).
CC   -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC   -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC       NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC   -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC       ACTIVATION OF NS3.
CC   -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC   -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC       ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC       AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC   -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC       PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC   -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC   -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M67463; AAA45534.1; -.
DR   PIR; A36814; GNWVCH.
DR   PDB; 1HEI; 25-NOV-98.
DR   PDB; 1A1V; 16-FEB-99.
DR   PDB; 1A1R; 17-JUN-98.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   TRANSFAC; T04155; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
```

```
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    191       CAPSID PROTEIN C.
FT   CHAIN       192    383       ENVELOPE GLYCOPROTEIN E1.
FT   CHAIN       384    746       ENVELOPE GLYCOPROTEIN E2.
FT   CHAIN       747    809       PROTEIN P7.
FT   CHAIN       810   1026       NONSTRUCTURAL PROTEIN NS2.
FT   CHAIN      1027   1657       PROTEASE/HELICASE NS3.
FT   CHAIN      1658   1711       NONSTRUCTURAL PROTEIN NS4A.
FT   CHAIN      1712   1972       NONSTRUCTURAL PROTEIN NS4B.
FT   CHAIN      1973   2420       NONSTRUCTURAL PROTEIN NS5A.
FT   CHAIN      2421   3011       NONSTRUCTURAL PROTEIN NS5B.
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
FT   SITE       1316   1319       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    476    476       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   STRAND     1224   1226
FT   TURN       1232   1233
FT   TURN       1236   1238
FT   HELIX      1239   1246
FT   TURN       1247   1248
FT   STRAND     1251   1255
FT   HELIX      1258   1271
FT   TURN       1272   1272
FT   STRAND     1277   1280
FT   TURN       1281   1282
FT   STRAND     1283   1285
FT   STRAND     1291   1295
FT   HELIX      1296   1301
FT   TURN       1302   1303
FT   STRAND     1312   1316
FT   TURN       1317   1319
FT   HELIX      1323   1335
FT   TURN       1336   1340
FT   STRAND     1343   1347
FT   TURN       1352   1353
FT   TURN       1360   1361
FT   STRAND     1362   1366
FT   STRAND     1368   1368
FT   STRAND     1373   1375
FT   TURN       1376   1377
FT   STRAND     1378   1380
FT   HELIX      1382   1385
FT   STRAND     1389   1393
FT   HELIX      1397   1409
FT   TURN       1410   1411
FT   STRAND     1414   1417
FT   TURN       1419   1420
FT   STRAND     1432   1436
FT   TURN       1438   1439
FT   STRAND     1450   1453
FT   STRAND     1456   1463
FT   STRAND     1471   1478
FT   STRAND     1480   1480
FT   HELIX      1481   1488
FT   TURN       1489   1490
FT   STRAND     1497   1501
FT   STRAND     1507   1507
FT   STRAND     1511   1511
FT   HELIX      1514   1527
FT   HELIX      1532   1544
FT   STRAND     1550   1550
FT   HELIX      1555   1564
FT   HELIX      1570   1578
FT   TURN       1579   1580
FT   HELIX      1584   1597
FT   TURN       1598   1598
FT   HELIX      1606   1611
FT   TURN       1614   1618
FT   STRAND     1622   1623
FT   STRAND     1627   1627
FT   STRAND     1635   1636
FT   HELIX      1640   1652
SQ   SEQUENCE   3011 AA;  327142 MW;  772CBB29CCD94753 CRC64;

  Query Match             87.2%;  Score 1385;  DB 1;  Length 3011;
  Best Local Similarity   83.8%;  Pred. No. 9.4e-112;
  Matches  254;  Conservative   26;  Mismatches   23;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              | : ||||||| |||:| | | || ||||||||| :|| ||||| ||:|| ||:|||| |
Db        904 ASLLKVPYFVRVQGLLRICALARKIAGGHYVQMAIIKLGALTGTCVYNHLAPLRDWAHNG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:|| || |:|||||||||||||:|||||||:||||||||    :||
Db        964 LRDLAVAVEPVVFSRMETKLITWGADTAACGDIINGLPVSARRGQEILLGPADGMVSKGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              |||||||||:|||||||||||||||||||||||||||:||||||:|||||:|||||||:|
Db       1024 RLLAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTATQTFLATCINGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              |||::|:| ||||: | |||||||||| || :||:|||||||||||||||||||||||
Db       1084 GAGTRTIASPKGPVIQTYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              |||||||||||||:|||||||||||||:|||||:|||||||||||||||||::||
Db       1144 RRGDSRGSLLSPRPISYLKGSSGGPLLCPTGHAVGLFRAAVCTRGVAKAVDFIPVENLET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 7
POLG_HCVJ6
ID   POLG_HCVJ6     STANDARD;      PRT;  3033 AA.
AC   P26660;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate HC-J6) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
```

```
OC   Hepacivirus.
OX   NCBI_TaxID=11113;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92044440; PubMed=1658196;
RA   Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
RA   Machida A., Miyakawa Y., Mayumi M.;
RT   "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
RT   from a human carrier: comparison with reported isolates for conserved
RT   and divergent regions.";
RL   J. Gen. Virol. 72:2697-2704(1991).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D00944; BAA00792.1; -.
DR   PIR; JQ1303; JQ1303.
DR   HSSP; P27958; 1HEI.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    733       NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT   CHAIN       734   1010       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1011   1619       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1620   1866       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1867   2017       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2018   3033       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1087   1087       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1111   1111       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1169   1169       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1234   1241       ATP (POTENTIAL).
FT   SITE       1320   1323       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    477    477       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    534    534       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    542    542       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    558    558       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    578    578       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    627    627       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    649    649       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1091   1091       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2038   2038       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2811   2811       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3033 AA;  329165 MW;  F957F5C1A273BE9E CRC64;

  Query Match             73.8%;  Score 1172;  DB 1;  Length 3033;
  Best Local Similarity   69.4%;  Pred. No. 3e-93;
  Matches  209;  Conservative   44;  Mismatches   48;  Indels    0;  Gaps    0;

Qy          3 ITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAGLR 62
              :|:|||||||  |:| | :||  ||| |||| : |   ||||:||||||: |||  |||
Db        910 LTRVPYFVRAHALLRMCTMVRHLAGGRYVQMVLLALGRWTGTYIYDHLTPMSDWAANGLR 969

Qy         63 DLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGWRL 122
              ||||||||:||| || |:| |||:||||||: ||||||| |||:|||||| :  :|| |
Db        970 DLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPADGYTSKGWSL 1029

Qy        123 LAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFHGA 182
              |||||||:|||||||| |: |:||||| :  ||:||:|| ||||| | ::|| |||:|||
Db       1030 LAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTTISGVLWTVYHGA 1089

Qy        183 GSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVRRR 242
              |:||||| :||:|||:: : ||||| :||| :|: |||||: |||||||:||||| |||
Db       1090 GNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRNADVIPARRR 1149

Qy        243 GDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTM 302
              || ||:||||||:| ||||||||:||| |||||:||||||:|||||::|||||||::::
Db       1150 GDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFIPVETLDIVT 1209

Qy        303 R 303
              |
Db       1210 R 1210


RESULT 8
POLG_HCVJ8
ID   POLG_HCVJ8     STANDARD;      PRT;  3033 AA.
AC   P26661;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
```

```
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate HC-J8) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11115;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92230232; PubMed=1314459;
RA   Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
RA   Fukuda S., Tsuda F., Mishiro S.;
RT   "Full-length sequence of a hepatitis C virus genome having poor
RT   homology to reported isolates: comparative study of four distinct
RT   genotypes.";
RL   Virology 188:331-341(1992).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10988; BAA01761.1; -.
DR   PIR; A40250; GNWVJ8.
DR   HSSP; P27958; 1HEI.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    733       NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT   CHAIN       734   1010       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1011   1619       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1620   1866       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1867   2017       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2018   3033       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1087   1087       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1111   1111       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1169   1169       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1234   1241       ATP (POTENTIAL).
FT   SITE       1320   1323       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    233    233       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    299    299       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    477    477       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    534    534       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    542    542       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    558    558       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    578    578       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    627    627       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    649    649       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1091   1091       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2038   2038       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2359   2359       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2811   2811       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3033 AA;  330177 MW;  1A173E7E3381FD1A CRC64;

  Query Match             72.4%;  Score 1150;  DB 1;  Length 3033;
  Best Local Similarity   67.3%;  Pred. No. 2.4e-91;
  Matches  204;  Conservative   47;  Mismatches   52;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              | : ::|||||  |:| | ||:  ||  |:||  : :   ||||:||||:||  || |
Db        908 ASLLRIPYFVRAHALLRVCTLVKHLAGARYIQMLLITIGRWTGTYIYDHLSPLSTWAAQG 967

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||:|||||:|| || |:| |||:| |||||: ||||||| |||:|||||| :  :||
Db        968 LRDLAIAVEPVVFSPMEKKVIVWGAETVACGDILHGLPVSARLGREVLLGPADGYTSKGW 1027

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              :|||||||:|||||||| |: ||||||||:  |:|||:|: ||:|| |.::|| |||:|
Db       1028 KLLAPITAYTQQTRGLLGAIVVSLTGRDKNEQAGQVQVLSSVTQTFLGTSISGVLWTVYH 1087

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              |||:|||||||||:|||||: : ||||| :||| :|: |||||: |||||||:|||||||
Db       1088 GAGNKTLAGPKGPVTQMYTSAEGDLVGWPSPPGTKSLDPCTCGAVDLYLVTRNADVIPVR 1147

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              |: | ||:||||||:| |||||||||:||  |||||:|||||| |||||::|||||||::
Db       1148 RKDDRRGALLSPRPLSTLKGSSGGPVLCSRGHAVGLFRAAVCARGVAKSIDFIPVESLDV 1207

Qy        301 TMR 303
                 |
Db       1208 ATR 1210


RESULT 9
DPO2_MOUSE
ID   DPO2_MOUSE     STANDARD;      PRT;   600 AA.
AC   P33611;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
```

```
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   DNA polymerase alpha 70 kDa subunit (DNA polymerase subunit B).
GN   POLA2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 84-102; 269-285 AND 394-403.
RX   MEDLINE=93216788; PubMed=8463324;
RA   Miyazawa H., Izumi M., Tada S., Takada R., Masutani M., Ui M.,
RA   Hanaoka F.;
RT   "Molecular cloning of the cDNAs for the four subunits of mouse DNA
RT   polymerase alpha-primase complex and their gene expression during
RT   cell proliferation and the cell cycle.";
RL   J. Biol. Chem. 268:8111-8122(1993).
CC   -!- FUNCTION: May play an essential role at the early stage of
CC       chromosomal DNA replication by coupling the polymerase
CC       alpha/primase complex to the cellular replication machinery (By
CC       similarity).
CC   -!- SUBUNIT: DNA polymerase alpha-primase is a four subunit enzyme
CC       (subunits A, B, C and D), which is assembled throughout the cell
CC       cycle. The largest subunit (subunit A) has DNA polymerase
CC       activity, the two smallest subunits (subunits C and D) have DNA
CC       primase activity. Subunit B binds to subunit A.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- PTM: PHOSPHORYLATED IN A CELL CYCLE-DEPENDENT MANNER, IN G2/M
CC       PHASE (BY SIMILARITY).
CC   -!- SIMILARITY: Belongs to the DNA polymerase alpha subunit B family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D13546; BAA02746.1; -.
DR   PIR; B46642; B46642.
DR   MGD; MGI:99690; Pola2.
DR   InterPro; IPR007200; DNA_pol_alpha_B.
DR   Pfam; PF04058; DNA_pol_alpha_B; 1.
KW   DNA replication; Nuclear protein; Phosphorylation.
FT   DOMAIN      101    107       POLY-GLU.
FT   DOMAIN      115    157       PRO/SER/THR-RICH (HYDROPHILIC).
SQ   SEQUENCE   600 AA;  66267 MW;  79F94BE6EF33FEBC CRC64;

  Query Match             6.4%;  Score 101;  DB 1;  Length 600;
  Best Local Similarity   24.8%;  Pred. No. 0.42;
  Matches   55;  Conservative   34;  Mismatches   71;  Indels   62;  Gaps   12;

Qy         15 LIRACMLVRKAAGGHYVQM-AFMKLAALT-------GTYVYDHL---------TPLQDWA 57
              |  |:| |:   |   :: ||  |  |        :: |: |         :  :|
Db         27 LAELCVLYRQTEDGMVSELIAFCTSAGKTCLTVDILNSFEYEVLNKKLSKAWHSASKDSG 86

Qy         58 HAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDI--ISGLP--------VSARRGREI 107
              ||| ||: |::: :| :: | : :      | : |  :|  |         :|||  |:
Db         87 HAGTRDI-VSIQELIEAEEEEETLLSSYTTPSKGPLKRVSSTPETPLTKRSVAARSPRQ- 144

Qy        108 LLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFL 167
              || |:           :|     ||:         || | |      |||     : |
Db        145 LLSPSS---------FSPSATPSQK--------YTSRTNR------GEVVTTFGSAQ--- 178

Qy        168 ATCVNGVCWTVFHGAGSKTL--AGPKGPITQMYTNVDQDLVG 207
                  |: |:   |:|| :|    |   |:|  |  : | |:|
Db        179 -----GLSWSGRGGSGSVSLKVVGDPEPLTGSYKAMFQQLMG 215


RESULT 10
VST2_HEVBU
ID   VST2_HEVBU     STANDARD;      PRT;   660 AA.
AC   P29326;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   Structural protein 2 precursor (ORF2).
OS   Hepatitis E virus (strain Burma) (HEV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage;
OC   Hepatitis E-like viruses.
OX   NCBI_TaxID=31767;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92024067; PubMed=1926770;
RA   Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,
RA   Fry K.E., Reyes G.R.;
RT   "Hepatitis E virus (HEV): molecular cloning and sequencing of the
RT   full-length viral genome.";
RL   Virology 185:120-131(1991).
CC   -!- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
CC       THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
CC       BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M73218; AAA45736.1; -.
DR   PIR; C40778; VHWWH2.
DR   InterPro; IPR004261; SP2.
DR   InterPro; IPR008975; Viral_cap_coat.
DR   Pfam; PF03014; SP2; 1.
KW   Signal.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    660       STRUCTURAL PROTEIN 2.
SQ   SEQUENCE   660 AA;  70978 MW;  5832A013CCC4A61C CRC64;

  Query Match             5.9%;  Score 93;  DB 1;  Length 660;
  Best Local Similarity   19.0%;  Pred. No. 2.3;
  Matches   72;  Conservative   47;  Mismatches  119;  Indels  140;  Gaps   16;

Qy         12 AQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPV 71
              |:  ||   ||  | ||: : ::|      | | |         |      |: : |:|
Db        188 ARATIRYRPLVPNAVGGYAISISFWPQTTTTPTSV--------DMNSITSTDVRILVQPG 239

Qy         72 IFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPAD--NFEGQGWRLLAPI-TA 128
              | |:: :                                  |::  ::  |||| :     |
Db        240 IASELVI---------------------------------PSERLHYRNQGWRSVETSGVA 267

Qy        129 YSQQTRGLL-----GCIITSLTG---------------------------------- 146
                : | ||:      | :: | |
Db        268 EEEATSGLVMLCIHGSLVNSYTNTPYTGALGLLDFALELEFRNLTPGNTNTRVSRYSSTA 327

Qy        147 --RDKNQVEGEVQVVSTATQSFLA----TCVNGV-------CWTVFH------------- 180
                | :   :|  :: :||   |:     |  |||         |:|:
Db        328 RHRLRRGADGTAELTTTAATRFMKDLYFTSTNGVGEIGRGIALTLFNLADTLLGGLPTEL 387

Qy        181 --GAG-----SKTLAGPKG-PITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLV-- 230
                ||      |: :    | |  ::||:|:      |  |              | | : :
Db        388 ISSAGGQLFYSRPVVSANGEPTVKLYTSVENA----QQDKGIAIPHDIDLGESRVVIQDY 443

Qy        231 -TRHADVIPVRRRGDSRG-SLLSPRPVSYLK----------GSSGGPLLCPSGHAVGIF 277
               :|     |       ||  |:|     | :|                ||| ||:     :| :
Db        444 DNQHEQDRPTPSPAPSRPFSVLRANDVLWLSLTAAEYDQSTYGSSTGPVYV--SDSVTLV 501

Qy        278 RAAVCTRGVAKAVDFIPV 295
                |   : ||:::|:  |
Db        502 NVATGAQAVARSLDWTKV 519


RESULT 11
VST2_HEVPA
```

```
ID   VST2_HEVPA     STANDARD;      PRT;   660 AA.
AC   P33426;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   Structural protein 2 precursor (ORF2).
OS   Hepatitis E virus (strain Pakistan) (HEV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage;
OC   Hepatitis E-like viruses.
OX   NCBI_TaxID=33774;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92115700; PubMed=1731327;
RA   Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,
RA   Malik I.A., Iqbal M., Purcell R.H.;
RT   "Characterization of a prototype strain of hepatitis E virus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).
CC   -!- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
CC       THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
CC       BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M80581; AAA45727.1; -.
DR   InterPro; IPR004261; SP2.
DR   InterPro; IPR008975; Viral_cap_coat.
DR   Pfam; PF03014; SP2; 1.
KW   Signal.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   CHAIN        23    660       STRUCTURAL PROTEIN 2.
SQ   SEQUENCE   660 AA;  70980 MW;  8085BC53CFB46FD3 CRC64;

  Query Match           5.9%;  Score 93;  DB 1;  Length 660;
  Best Local Similarity  19.0%;  Pred. No. 2.3;
  Matches  72;  Conservative  49;  Mismatches  117;  Indels  140;  Gaps  17;

Qy         12 AQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPV 71
              |:  ||   ||  | ||: : ::|      | | |        |       |: : |:|
Db        188 ARATIRYRPLVPNAVGGYAISISFWPQTTTTPTSV--------DMNSITSTDVRILVQPG 239

Qy         72 IFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPAD--NFEGQGWRLL------ 123
              | |:: :                              |::  ::  |||| :
Db        240 IASELVI--------------------------------PSERLHYRNQGWRSVETSGVA 267

Qy        124 ---------------APITAYSQQT-RGLLGCI-------ITSLTGRDKNQ--------- 151
                           :|: :|:     | || :        :||  : |
Db        268 EEEATSGLVMLCIHGSPVNSYTNTPYTGALGLLDFALELEFRNLTPGNTNTRVSRYSSTA 327

Qy        152 -------VEGEVQVVSTATQSFLA----TCVNGV-------CWTVFH------------- 180
                  :|  :: :||   |:     |  |||          |:|:
Db        328 RHRLRRGADGTAELTTTAATRFMKDLYFTSTNGVGEIGRGIALTLFNLADTLLGGLPTEL 387

Qy        181 --GAG-----SKTLAGPKG-PITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLV-- 230
                ||     |: :    | |  ::||:|:      |   |         | | : :
Db        388 ISSAGGQLFYSRPVVSANGEPTVKLYTSVENA----QQDKGIAIPHDIDLGESRVVIQDY 443

Qy        231 -TRHADVIPVRRRGDSRG-SLLSPRPVSYLK----------GSSGGPLLCPSGHAVGIF 277
               :|    |       ||  |:|    | :|          ||| ||:      :| :
Db        444 DNQHEQDRPTPSPAPSRPFSVLRANDVLWLSLTAAEYDQSTYGSSTGPVYV--SDSVTLV 501

Qy        278 RAAVCTRGVAKAVDFIPV 295
               |   : ||:::|:  |
Db        502 NVATGAQAVARSLDWTKV 519


RESULT 12
TRFE_HORSE
```

```
ID   TRFE_HORSE     STANDARD;      PRT;   706 AA.
AC   P27425;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Serotransferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal
DE   binding globulin).
GN   TF.
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93277958; PubMed=8504171;
RA   Carpenter M.A., Broad T.E.;
RT   "The cDNA sequence of horse transferrin.";
RL   Biochim. Biophys. Acta 1173:230-232(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Extraembryonic tissue;
RA   McDowell K.J., Adams M.H., Baker C.B.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Transferrins are iron binding transport proteins which
CC       can bind two atoms of ferric iron in association with the binding
CC       of an anion, usually bicarbonate. It is responsible for the
CC       transport of iron from sites of absorption and heme degradation to
CC       those of storage and utilization. Serum transferrin may also have
CC       a further role in stimulating cell proliferation.
CC   -!- SUBUNIT: Monomer.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.
CC   -!- DOMAIN: Composed of two homologous domains.
CC   -!- SIMILARITY: Belongs to the transferrin family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M69020; AAA30958.1; -.
DR   EMBL; U21127; AAA63684.1; -.
DR   PIR; S33761; S33761.
DR   HSSP; P02787; 1A8E.
DR   InterPro; IPR001156; Transferrin.
DR   Pfam; PF00405; transferrin; 2.
DR   PRINTS; PR00422; TRANSFERRIN.
DR   SMART; SM00094; TR_FER; 2.
DR   PROSITE; PS00205; TRANSFERRIN_1; 2.
DR   PROSITE; PS00206; TRANSFERRIN_2; 2.
DR   PROSITE; PS00207; TRANSFERRIN_3; 2.
KW   Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
KW   Signal.
FT   SIGNAL        1     19       BY SIMILARITY.
FT   CHAIN        20    706       SEROTRANSFERRIN.
FT   REPEAT       20    357       1.
FT   REPEAT      358    706       2.
FT   DISULFID     26     64       BY SIMILARITY.
FT   DISULFID     36     55       BY SIMILARITY.
FT   DISULFID    134    215       BY SIMILARITY.
FT   DISULFID    174    190       BY SIMILARITY.
FT   DISULFID    177    198       BY SIMILARITY.
FT   DISULFID    187    200       BY SIMILARITY.
FT   DISULFID    248    262       BY SIMILARITY.
FT   DISULFID    360    623       BY SIMILARITY.
FT   DISULFID    366    398       BY SIMILARITY.
FT   DISULFID    376    389       BY SIMILARITY.
FT   DISULFID    423    701       BY SIMILARITY.
FT   DISULFID    441    664       BY SIMILARITY.
FT   DISULFID    474    550       BY SIMILARITY.
```

```
FT   DISULFID    498    692       BY SIMILARITY.
FT   DISULFID    508    522       BY SIMILARITY.
FT   DISULFID    519    533       BY SIMILARITY.
FT   DISULFID    590    604       BY SIMILARITY.
FT   DISULFID    642    647       BY SIMILARITY.
FT   METAL        79     79       IRON 1 (BY SIMILARITY).
FT   METAL       111    111       IRON 1 (BY SIMILARITY).
FT   METAL       209    209       IRON 1 (BY SIMILARITY).
FT   METAL       270    270       IRON 1 (BY SIMILARITY).
FT   METAL       413    413       IRON 2 (BY SIMILARITY).
FT   METAL       449    449       IRON 2 (BY SIMILARITY).
FT   METAL       544    544       IRON 2 (BY SIMILARITY).
FT   METAL       612    612       IRON 2 (BY SIMILARITY).
FT   BINDING     136    136       CARBONATE 1 (BY SIMILARITY).
FT   BINDING     140    140       CARBONATE 1 (BY SIMILARITY).
FT   BINDING     142    142       CARBONATE 1 (VIA AMIDE NITROGEN) (BY
FT                                SIMILARITY).
FT   BINDING     143    143       CARBONATE 1 (VIA AMIDE NITROGEN) (BY
FT                                SIMILARITY).
FT   BINDING     476    476       CARBONATE 2 (BY SIMILARITY).
FT   BINDING     480    480       CARBONATE 2 (BY SIMILARITY).
FT   BINDING     482    482       CARBONATE 2 (VIA AMIDE NITROGEN) (BY
FT                                SIMILARITY).
FT   BINDING     483    483       CARBONATE 2 (VIA AMIDE NITROGEN) (BY
FT                                SIMILARITY).
FT   CARBOHYD    515    515       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   706 AA;  78094 MW;  1A0FA566C0409D8A CRC64;

  Query Match             5.8%;  Score 92.5;  DB 1;  Length 706;
  Best Local Similarity   21.5%;  Pred. No. 2.8;
  Matches   64;  Conservative   43;  Mismatches  109;  Indels   81;  Gaps   17;

Qy         33 MAFMKLAALTGTYV---YDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGA---- 85
              : |:::     |::   |:::|         :|:|   :  |  :  | | | : | |
Db        321 LGFLRIPPAMDTWLYLGYEYVT--------AIRNLREDIRPEVPKD-ECKKVKWCAIGHH 371

Qy         86 DTAACGD--IISGLPVSARRGR-------EILLGPADNFEGQGWRL-------LAPITAY 129
              :   | :  : || :     :       :|: | ||      |  :       | |: |
Db        372 EKVKCDEWSVNSGGNIECESAQSTEDCIAKIVKGEADAMSLDGGFIYIAGKCGLVPVLAE 431

Qy        130 SQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAG 189
              : :||    |: |   |       | || :::   |          |          :| |
Db        432 NYETRSGSACVDTPEEGYH------AVAVVKSSSDPDLT-------W--------NSLKG 470

Qy        190 PKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSL 249
              |      :| ||:   ||  | |           ||   :| :     | | : |
Db        471 KK----SCHTGVDR-TAGWNIPMGL-----------LYSEIKHCEFDKFFREGCAPGYR 513

Qy        250 LSPRPVSYLKGSSGGP-LLC-PSGHA-----VGIFRAAVCTRGVAKAVDFIPVESME 299
              :    :   ||: ||   | |: |       | ||  |    ||    |:  :::|
Db        514 RNSTLCNLCIGSASGPGRECEPNNHERYYGYTGAFRCLVEKGDVA----FVKHQTVE 566


RESULT 13
VST2_HEVME
ID   VST2_HEVME     STANDARD;      PRT;   659 AA.
AC   Q03500;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   Structural protein 2 precursor.
OS   Hepatitis E virus (strain Mexico) (HEV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage;
OC   Hepatitis E-like viruses.
OX   NCBI_TaxID=31768;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93079857; PubMed=1448913;
RA   Huang C.C., Nguyen D., Fernandez J., Yun K.Y., Fry K.E.,
RA   Bradley D.W., Tam A.W., Reyes G.R.;
RT   "Molecular cloning and sequencing of the Mexico isolate of hepatitis
RT   E virus (HEV).";
RL   Virology 191:550-558(1992).

CC   -!- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
CC       THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
CC       BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M74506; AAA45732.1; -.
DR   PIR; B44212; B44212.
DR   InterPro; IPR004261; SP2.
DR   InterPro; IPR008975; Viral_cap_coat.
DR   Pfam; PF03014; SP2; 1.
KW   Signal.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   CHAIN        23    659       STRUCTURAL PROTEIN 2.
SQ   SEQUENCE   659 AA;  70640 MW;  CF75E75EFD8FBE2C CRC64;

  Query Match             5.6%;  Score 89.5;  DB 1;  Length 659;
  Best Local Similarity   17.8%;  Pred. No. 4.7;
  Matches   67;  Conservative   50;  Mismatches  121;  Indels  139;  Gaps   14;

Qy         12 AQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPV 71
              |:  ||   ||  | ||: : ::|      | | |       |       |: : |:|
Db        188 ARATIRYRPLVPNAVGGYAISISFWPQTTTTPTSV--------DMNSITSTDVRILVQPG 239

Qy         72 IFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPAD--NFEGQGWRLL------ 123
              | |:: :                          |::  ::  |||| :
Db        240 IASELVI--------------------------------PSERLHYRNQGWRSVETSGVA 267

Qy        124 ---------------APITAYSQ---------------QTRGLLGCIITSLTGR----- 147
                            :|: :|:                : | |  |   :   |
Db        268 EEEATSGLVMLCIHGSPVNSYTNTPYTGALGLLDFALELEFRNLTTCNTNTRVSRYSSTA 327

Qy        148 --DKNQVEGEVQVVSTATQSFLA----TCVNGV-------------CWTVFHGAGSKTL 187
                   :|  :: :||   |:     | :|||                |:  |  :: :
Db        328 RHSARGADGTAELTTTAATRFMKDLHFTGLNGVGEVGRGIALTLLNLADTLLGGLPTELI 387

Qy        188 AGPKG-------------PITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLV--- 230
              :  |             | ::||:|:     |  |          | | : :
Db        388 SSAGGQLFYSRPVVSANGEPTVKLYTSVENA----QQDKGVAIPHDIDLGDSRVVIQDYD 443

Qy        231 TRHADVIPVRRRGDSRG-SLLSPRPVSYLK----------GSSGGPLLCPSGHAVGIFR 278
              :|    |      ||  |:|    | :|           |||  ||:      :| :
Db        444 NQHEQDRPTPSPAPSRPFSVLRANDVLWLSLTAAEYDQSTYGSSTGPVYI--SDSVTLVN 501

Qy        279 AAVCTRGVAKAVDFIPV 295
              |   : ||:::|:  |
Db        502 VATGAQAVARSLDWSKV 518


RESULT 14
POLG_TBEVW
ID   POLG_TBEVW     STANDARD;      PRT;  3414 AA.
AC   P14336; Q88493;
DT   01-JAN-1990 (Rel. 13, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
DE   protein (Envelope protein M); Major envelope protein E; Nonstructural
DE   proteins NS1, NS2A, NS2B, NS4A and NS4B; Protease/helicase
DE   (EC 3.4.21.98) (NS3); RNA-directed RNA polymerase (EC 2.7.7.48)
DE   (NS5)].
OS   Tick-borne encephalitis virus (Western subtype) (TBEV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Flavivirus.
OX   NCBI_TaxID=11088;
RN   [1]
RP   SEQUENCE FROM N.A., AND REVISIONS.
```

```
RC   STRAIN=Neudoerfl;
RX   MEDLINE=96036491; PubMed=7483260;
RA   Wallner G., Mandl C.W., Kunz C., Heinz F.X.;
RT   "The flavivirus 3'-noncoding region: extensive size heterogeneity
RT   independent of evolutionary relationships among strains of tick-borne
RT   encephalitis virus.";
RL   Virology 213:169-178(1995).
RN   [2]
RP   SEQUENCE OF 1-779 FROM N.A.
RC   STRAIN=Neudoerfl;
RX   MEDLINE=88322870; PubMed=3413985;
RA   Mandl C.W., Heinz F.X., Kunz C.;
RT   "Sequence of the structural proteins of tick-borne encephalitis virus
RT   (western subtype) and comparative analysis with other flaviviruses.";
RL   Virology 166:197-205(1988).
RN   [3]
RP   SEQUENCE OF 767-3414 FROM N.A.
RC   STRAIN=Neudoerfl;
RX   MEDLINE=90051080; PubMed=2554575;
RA   Mandl C.W., Heinz F.X., Stoeckl E., Kunz C.;
RT   "Genome sequence of tick-borne encephalitis virus (Western subtype)
RT   and comparative analysis of nonstructural proteins with other
RT   flaviviruses.";
RL   Virology 173:291-301(1989).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 281-680.
RX   MEDLINE=95272700; PubMed=7753193;
RA   Rey F.A., Heinz F.X., Mandl C.W., Kunz C., Harrison S.C.;
RT   "The envelope glycoprotein from tick-borne encephalitis virus at 2-A
RT   resolution.";
RL   Nature 375:291-298(1995).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function.
CC       NS3 and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U27495; AAA86870.1; -.
DR   PIR; A31052; GNWVNE.
DR   PDB; 1SVB; 10-JUN-96.
DR   MEROPS; S07.UPW; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001122; Flavi_capsidC.
DR   InterPro; IPR000336; Flavi_glycoprotE.
DR   InterPro; IPR000069; Flavi_M.
DR   InterPro; IPR001157; Flavi_NS1.
DR   InterPro; IPR000752; Flavi_NS2A.
DR   InterPro; IPR000487; Flavi_NS2B.
DR   InterPro; IPR000404; Flavi_NS4A.
DR   InterPro; IPR001528; Flavi_NS4B.
DR   InterPro; IPR000208; Flavi_NS5.
DR   InterPro; IPR002535; Flavi_propep.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001850; Peptidase_S7.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   InterPro; IPR002877; RrmJ_FtsJ.
DR   Pfam; PF01003; Flavi_capsid; 1.
DR   Pfam; PF02832; Flavi_glycop_C; 1.
DR   Pfam; PF00869; Flavi_glycoprot; 1.
DR   Pfam; PF00949; Flavi_helicase; 1.
DR   Pfam; PF01004; Flavi_M; 1.
DR   Pfam; PF00948; Flavi_NS1; 1.
DR   Pfam; PF01005; Flavi_NS2A; 1.
DR   Pfam; PF01002; Flavi_NS2B; 1.
DR   Pfam; PF01350; Flavi_NS4A; 1.
DR   Pfam; PF01349; Flavi_NS4B; 1.
DR   Pfam; PF00972; Flavi_NS5; 1.
DR   Pfam; PF01570; Flavi_propep; 1.
DR   Pfam; PF01728; FtsJ; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   ProDom; PD001556; Flavi_glycoprotE; 1.
DR   ProDom; PD001496; Flavi_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00690; DEAH_ATP_HELICASE; FALSE_NEG.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Hydrolase; Helicase;
KW   ATP-binding; Transmembrane; Nonstructural protein; 3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    112       CAPSID PROTEIN C.
FT   PROPEP      113    205
FT   CHAIN       206    280       ENVELOPE GLYCOPROTEIN M.
FT   CHAIN       281    776       MAJOR ENVELOPE PROTEIN E.
FT   CHAIN       777   1128       NONSTRUCTURAL PROTEIN NS1.
FT   CHAIN      1129   1358       NONSTRUCTURAL PROTEIN NS2A.
FT   CHAIN      1359   1489       NONSTRUCTURAL PROTEIN NS2B.
FT   CHAIN      1490   2110       PROTEASE/HELICASE (NS3).
FT   CHAIN      2111   2259       NONSTRUCTURAL PROTEIN NS4A.
FT   CHAIN      2260   2511       NONSTRUCTURAL PROTEIN NS4B.
FT   CHAIN      2512   3414       RNA-DIRECTED RNA POLYMERASE (NS5).
FT   DOMAIN      378    391       INVOLVED IN FUSION.
FT   NP_BIND    1688   1695       ATP (POTENTIAL).
FT   SITE       1779   1782       DEAH BOX.
FT   TRANSMEM    101    112       HYDROPHOBIC SIGNAL SEQUENCE (POTENTIAL).
FT   TRANSMEM    247    259       POTENTIAL.
FT   TRANSMEM    266    280       POTENTIAL.
FT   TRANSMEM    738    751       POTENTIAL.
FT   DISULFID    283    310
FT   DISULFID    340    396
FT   DISULFID    354    385
FT   DISULFID    372    401
FT   DISULFID    466    570
FT   DISULFID    587    618
FT   CARBOHYD    144    144       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    434    434       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    861    861       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    983    983       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    999    999       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2447   2447       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   HELIX       282    285
FT   STRAND      290    294
FT   STRAND      300    306
FT   TURN        307    308
FT   STRAND      311    315
FT   TURN        316    317
FT   STRAND      318    331
FT   STRAND      334    340
FT   STRAND      343    352
FT   TURN        355    356
FT   HELIX       363    366
FT   TURN        368    369
FT   STRAND      370    379
FT   HELIX       381    383
FT   TURN        384    384
FT   STRAND      389    401
FT   TURN        403    404
FT   STRAND      406    411
```

```
FT   TURN        414    416
FT   STRAND      418    425
FT   TURN        433    434
FT   TURN        438    439
FT   STRAND      441    446
FT   TURN        447    448
FT   STRAND      449    449
FT   STRAND      451    455
FT   HELIX       457    459
FT   STRAND      460    467
FT   HELIX       468    470
FT   TURN        475    476
FT   STRAND      477    481
FT   TURN        484    485
FT   TURN        487    488
FT   STRAND      492    496
FT   HELIX       497    501
FT   TURN        502    502
FT   STRAND      507    508
FT   TURN        510    511
FT   STRAND      516    516
FT   TURN        517    517
FT   HELIX       518    521
FT   STRAND      522    524
FT   STRAND      528    528
FT   TURN        529    530
FT   STRAND      531    531
FT   STRAND      534    536
FT   HELIX       541    547
FT   TURN        548    550
FT   STRAND      553    557
FT   TURN        558    559
FT   STRAND      560    562
FT   STRAND      567    573
FT   TURN        575    576
FT   STRAND      577    577
FT   TURN        582    583
FT   STRAND      586    586

  Query Match             5.5%;  Score 88;  DB 1;  Length 3414;
  Best Local Similarity   23.0%;  Pred. No. 47;
  Matches   59;  Conservative   28;  Mismatches   86;  Indels   84;  Gaps   12;

Qy         49 HLTPLQD-------WAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSA 101
              ||| |:       |  |||     |   : :| :   :  |           | | |
Db       1442 HLTELEKEERMMAFWLIAGL-----AASAIHWSGILGVMGLWTLTEMLRSSRRSDLVFSG 1496

Qy        102 RRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVST 161
              : |||    | :  :|  :|: :|   : |                ||| |
Db       1497 QGGRERGDRPFEVKDGV-YRIFSPGLFWGQ-----------------NQVG-----VGY 1532

Qy        162 ATQSFLATCVNGVCWTVFHGAG---SKTLAGPKGPITQMYTNVDQDLV----------GW 208
              ::  | |      | |  ||       :|||      : :| :|:|           |
Db       1533 GSKGVLHT-----MWHVTRGAALSIDDAVAGP------YWADVREDVVCYGGAWSLEEKW 1581

Qy        209 QA--------PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKG 260
              :          ||| |:     |   :| | |               |  |    |:  :||
Db       1582 KGETVQVHAFPPG-RAHEVHQCQPGELILDT--------------GRKLGAIPIDLVKG 1625

Qy        261 SSGGPLLCPSGHAVGIF 277
              :|| |:|    |  ||::
Db       1626 TSGSPILNAQGVVVGLY 1642


RESULT 15
POLG_TBEVS
ID   POLG_TBEVS     STANDARD;      PRT;  3412 AA.
AC   P07720; P07721; Q88475; Q88476; Q88477; Q88478; Q88479; Q88877;
AC   Q88878; Q88879;
DT   01-APR-1988 (Rel. 07, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
DE   protein (Envelope protein M); Major envelope protein E; Nonstructural
DE   proteins NS1, NS2A, NS2B, NS4A and NS4B; Protease/helicase
DE   (EC 3.4.21.98) (NS3); RNA-directed RNA polymerase (EC 2.7.7.48)
DE   (NS5)].
OS   Tick-borne encephalitis virus (strain Sofjin) (TBEV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Flavivirus.
OX   NCBI_TaxID=11087;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90101381; PubMed=2136778;
RA   Pletnev A.G., Yamshchikov V.F., Blinov V.M.;
RT   "Nucleotide sequence of the genome and complete amino acid sequence
RT   of the polyprotein of tick-borne encephalitis virus.";
RL   Virology 174:250-263(1990).
RN   [2]
RP   SEQUENCE OF 1-1190 FROM N.A.
RX   MEDLINE=88319988; PubMed=2970626;
RA   Yamshchikov V.F., Pletnev A.G.;
RT   "Nucleotide sequence of the genome region encoding the structural
RT   proteins and the NS1 protein of the tick borne encephalitis virus.";
RL   Nucleic Acids Res. 16:7750-7750(1988).
RN   [3]
RP   SEQUENCE OF 1-683 AND 777-1002 FROM N.A.
RX   MEDLINE=86220766; PubMed=3709796;
RA   Pletnev A.G., Yamshchikov V.F., Blinov V.M.;
RT   "Tick-borne encephalitis virus genome. The nucleotide sequence coding
RT   for virion structural proteins.";
RL   FEBS Lett. 200:317-321(1986).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function.
CC       NS3 and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- MISCELLANEOUS: The nonstructural protein NS1 presents two
CC       alternative cleavage sites for its C-terminus, which may define a
CC       soluble or a membrane-bound form of NS1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X07755; CAA30581.1; -.
DR   EMBL; X03870; CAA27500.1; -.
DR   EMBL; X03870; CAA27501.1; ALT_SEQ.
DR   EMBL; X03870; CAA27502.1; ALT_SEQ.
DR   EMBL; X03870; CAA27503.1; ALT_SEQ.
DR   EMBL; X03871; CAA27505.1; -.
DR   MEROPS; S07.UPW; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001122; Flavi_capsidC.
DR   InterPro; IPR000336; Flavi_glycoprotE.
DR   InterPro; IPR000069; Flavi_M.
DR   InterPro; IPR001157; Flavi_NS1.
DR   InterPro; IPR000752; Flavi_NS2A.
DR   InterPro; IPR000487; Flavi_NS2B.
DR   InterPro; IPR000404; Flavi_NS4A.
DR   InterPro; IPR001528; Flavi_NS4B.
DR   InterPro; IPR000208; Flavi_NS5.
DR   InterPro; IPR002535; Flavi_propep.
DR   InterPro; IPR001650; Helicase_C.
```

```
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001850; Peptidase_S7.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   InterPro; IPR002877; RrmJ_FtsJ.
DR   Pfam; PF01003; Flavi_capsid; 1.
DR   Pfam; PF02832; Flavi_glycop_C; 1.
DR   Pfam; PF00869; Flavi_glycoprot; 1.
DR   Pfam; PF00949; Flavi_helicase; 1.
DR   Pfam; PF01004; Flavi_M; 1.
DR   Pfam; PF00948; Flavi_NS1; 1.
DR   Pfam; PF01005; Flavi_NS2A; 1.
DR   Pfam; PF01002; Flavi_NS2B; 1.
DR   Pfam; PF01350; Flavi_NS4A; 1.
DR   Pfam; PF01349; Flavi_NS4B; 1.
DR   Pfam; PF00972; Flavi_NS5; 1.
DR   Pfam; PF01570; Flavi_propep; 1.
DR   Pfam; PF01728; FtsJ; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   ProDom; PD001556; Flavi_glycoprotE; 1.
DR   ProDom; PD001496; Flavi_NS1; 1.
DR   PROSITE; PS00690; DEAH_ATP_HELICASE; FALSE_NEG.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Hydrolase; Helicase;
KW   ATP-binding; Transmembrane; Nonstructural protein.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    112       CAPSID PROTEIN C.
FT   PROPEP      113    205
FT   CHAIN       206    280       ENVELOPE GLYCOPROTEIN M.
FT   CHAIN       281    776       MAJOR ENVELOPE PROTEIN E.
FT   CHAIN       777  ?1128       NONSTRUCTURAL PROTEIN NS1 (OR 1190).
FT   CHAIN     ?1129   1358       NONSTRUCTURAL PROTEIN NS2A (OR 1191).
FT   CHAIN      1359   1489       NONSTRUCTURAL PROTEIN NS2B.
FT   CHAIN      1490   2110       PROTEASE/HELICASE (NS3).
FT   CHAIN      2111   2259       NONSTRUCTURAL PROTEIN NS4A.
FT   CHAIN      2260   2510       NONSTRUCTURAL PROTEIN NS4B.
FT   CHAIN      2511   3412       RNA-DIRECTED RNA POLYMERASE (NS5).
FT   NP_BIND    1688   1695       ATP (POTENTIAL).
FT   SITE       1779   1782       DEAH BOX.
FT   TRANSMEM    101    112       HYDROPHOBIC SIGNAL SEQUENCE (POTENTIAL).
FT   TRANSMEM    247    259       POTENTIAL.
FT   TRANSMEM    266    280       POTENTIAL.
FT   TRANSMEM    738    751       POTENTIAL.
FT   DISULFID    283    310       BY SIMILARITY.
FT   DISULFID    340    396       BY SIMILARITY.
FT   DISULFID    354    385       BY SIMILARITY.
FT   DISULFID    372    401       BY SIMILARITY.
FT   DISULFID    466    570       BY SIMILARITY.
FT   DISULFID    587    618       BY SIMILARITY.
FT   CARBOHYD    144    144       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    434    434       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    861    861       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    983    983       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    999    999       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1228   1228       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2447   2447       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2466   2466       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    381    381       W -> S (IN REF. 3).
FT   CONFLICT    850    850       E -> D (IN REF. 3).
SQ   SEQUENCE   3412 AA;  377976 MW;  0F61CE6DCCDC5965 CRC64;

  Query Match             5.5%;  Score 87;  DB 1;  Length 3412;
  Best Local Similarity   24.4%;  Pred. No. 58;
  Matches   50;  Conservative   23;  Mismatches   58;  Indels   74;  Gaps   11;

Qy         95 SGLPVSARRGREILLGPADNPEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEG 154
              |||   | : |||     | :  :|  :|: :|          |||                |
Db       1490 SGLVFSGQGGRERGDRPFEVKDGV-YRIFSP---------GLLW---------------G 1524

Qy        155 EVQV-VSTATQSFLATCVNGVCWTVFHGAG---SKTLAGPKGPITQMYTNVDQDLV---- 206
              : || |   ::  | |       | |  ||       :|||       : :| :|:|
Db       1525 QRQVGVGYGSKGVLHT-----MWHVTRGAALSIDDAVAGP------YWADVKEDVVCYGG 1573

Qy        207 ------GWQA--------PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSP 252
                     |:         ||| |:     |   :| | |              |  :
Db       1574 AWSLEEKWKGETVQVHAFPPG-RAHEVHQCQPGELLLDT--------------GRRIGA 1617

Qy        253 RPVSYLKGSSGGPLLCPSGHAVGIF 277
                |:   ||:|| |:|   |  ||::
Db       1618 VPIDLAKGTSGSPILNSQGVVVGLY 1642


Search completed: May  6, 2004, 09:31:49
Job time : 7.32568 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:21:36 ; Search time 28.339 Seconds
(without alignments)
3373.509 Million cell updates/sec

Title: US-10-650-585-10
Perfect score: 1589
Sequence: 1 AGITKVPYFVRAQGLIRACM.....RGVAKAVDFIPVESMETTMR 303

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1550	97.5	3010	12	Q9J3H7	Q9j3h7 hepatitis c
2	1545	97.2	3010	12	Q68826	Q68826 hepatitis c
3	1545	97.2	3010	12	P90191	P90191 hepatitis c
4	1544	97.2	3010	12	Q9DTE6	Q9dte6 hepatitis c
5	1542	97.0	3010	12	Q9DTE4	Q9dte4 hepatitis c
6	1541	97.0	3010	12	Q9DTD6	Q9dtd6 hepatitis c
7	1540	96.9	3010	12	P88803	P88803 hepatitis c
8	1539	96.9	3010	12	Q9J3H5	Q9j3h5 hepatitis c
9	1539	96.9	3010	12	Q807P3	Q807p3 hepatitis c
10	1537	96.7	3010	12	Q9J3F9	Q9j3f9 hepatitis c
11	1534	96.5	361	12	O70815	O70815 hepatitis c
12	1534	96.5	3008	12	Q9J3F4	Q9j3f4 hepatitis c
13	1534	96.5	3010	12	Q9J3H3	Q9j3h3 hepatitis c
14	1533	96.5	3010	12	Q9J3H2	Q9j3h2 hepatitis c
15	1533	96.5	3010	12	Q9J3I0	Q9j3i0 hepatitis c
16	1532	96.4	3010	12	Q9QIY3	Q9qiy3 hepatitis c
17	1532	96.4	3010	12	O09796	O09796 hepatitis c
18	1531	96.3	3010	12	Q9QIY7	Q9qiy7 hepatitis c
19	1531	96.3	3010	12	Q9QIY8	Q9qiy8 hepatitis c
20	1531	96.3	3010	12	Q9QIX6	Q9qix6 hepatitis c
21	1531	96.3	3010	12	Q9QIX5	Q9qix5 hepatitis c
22	1531	96.3	3010	12	Q9J3H6	Q9j3h6 hepatitis c
23	1530	96.3	3010	12	Q91AU0	Q91au0 hepatitis c
24	1530	96.3	3010	12	Q9J3H9	Q9j3h9 hepatitis c
25	1530	96.3	3010	12	Q9WMX2	Q9wmx2 hepatitis c
26	1529	96.2	3013	12	Q9J3H4	Q9j3h4 hepatitis c
27	1528	96.2	1186	12	Q81755	Q81755 hepatitis c
28	1528	96.2	2284	12	Q81817	Q81817 hepatitis c
29	1528	96.2	3010	12	Q68788	Q68788 hepatitis c
30	1528	96.2	3010	12	P89966	P89966 hepatitis c
31	1528	96.2	3010	12	Q9DTD7	Q9dtd7 hepatitis c
32	1528	96.2	3010	12	Q9DTF0	Q9dtf0 hepatitis c
33	1528	96.2	3014	12	Q9DTE0	Q9dte0 hepatitis c
34	1527	96.1	3010	12	Q99AU2	Q99au2 hepatitis c
35	1527	96.1	3011	12	Q9DTE3	Q9dte3 hepatitis c
36	1526	96.0	361	12	O70818	O70818 hepatitis c
37	1526	96.0	3010	12	Q9QP61	Q9qp61 hepatitis c
38	1526	96.0	3010	12	Q9J3G6	Q9j3g6 hepatitis c
39	1525	96.0	3010	12	Q8QRL8	Q8qrl8 hepatitis c
40	1524	95.9	3010	12	Q9J3I1	Q9j3i1 hepatitis c
41	1523	95.8	3010	12	Q9QIY5	Q9qiy5 hepatitis c
42	1523	95.8	3010	12	Q9DTE9	Q9dte9 hepatitis c
43	1523	95.8	3010	12	Q9QIY4	Q9qiy4 hepatitis c
44	1523	95.8	3010	12	Q81541	Q81541 hepatitis c
45	1522	95.8	3010	12	Q9J3G3	Q9j3g3 hepatitis c

ALIGNMENTS

RESULT 1
Q9J3H7
ID Q9J3H7 PRELIMINARY; PRT; 3010 AA.
AC Q9J3H7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD15;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
RT progression.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF207756; AAF65946.1; -.
DR PIR; A61196; A61196.
DR PIR; PQ0246; PQ0246.
DR PIR; PQ0804; PQ0804.
DR PIR; PS0329; PS0329.
DR HSSP; P26663; 1JXP.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.

```
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327365 MW;  D8653F7317FFA106 CRC64;

  Query Match             97.5%;  Score 1550;  DB 12;  Length 3010;
  Best Local Similarity   95.7%;  Pred. No. 2.2e-127;
  Matches  290;  Conservative    9;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||::|||||||||||||||||||||:|||||||||||||:|||| ||| |||||
Db        904 AGITRMPYFVRAQGLIRACMLVRKAAGGHYIQMAFMKLAALTGTYIYDHLAPLQHWAHAG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:|||| |||||||||||||||| ||||||||||:|||||:||||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREVLLGPADSFEGQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCINGVCWTVFH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 2
Q68826
ID   Q68826      PRELIMINARY;      PRT;  3010 AA.
AC   Q68826;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J33;
RA   Cho M.J.;
RT   "Molecular cloning of Hepatitis C virus genome from a single Japanese
RT   patient.";
RL   Submitted (SEP-1991) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; D14484; BAA03375.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PQ0804; PQ0804.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW   Hydrolase; Nonstructural protein; Polyprotein;
KW   RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327150 MW;  7270F47984554FAD CRC64;

  Query Match             97.2%;  Score 1545;  DB 12;  Length 3010;
  Best Local Similarity   96.4%;  Pred. No. 6.1e-127;
```

```
  Matches  292;  Conservative    5;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              |||| |||||||||||||||||||| |||||||||||||||||||||||||:||||||
Db        904 AGITAVPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLRDWAHAG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:||||| |||||||||||||||| ||||||||||||||||: ||||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGPADSLEGQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              |||||||||||||||||||||||||||||||||||||||||||||||||:|||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCINGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 RRGDGRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 3
P90191
ID   P90191      PRELIMINARY;      PRT;  3010 AA.
AC   P90191;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCV-1b;
RA   Enomoto N.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCV-1b;
RX   MEDLINE=95340824; PubMed=7542279;
RA   Enomoto N., Sakuma I., Asahina Y., Kurosaki M., Murakami T.,
RA   Yamamoto C., Izumi N., Marumo F., Sato C.;
RT   "Comparison of full-length sequences of interferon-sensitive and
RT   resistant hepatitis C virus 1b.";
RL   J. Clin. Invest. 96:224-230(1995).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; D50482; BAA09073.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0254; PQ0254.
DR   PIR; PQ0804; PQ0804.
DR   PIR; PS0329; PS0329.
DR   PDB; 1DXW; 12-JAN-01.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
FT   CHAIN         1    191       CORE PROTEIN.
FT   CHAIN       192    383       E1.
FT   CHAIN       384    809       E2.
FT   CHAIN       810   1026       NS2.
FT   CHAIN      1027   1657       NS3.
FT   CHAIN      1658   1711       NS4A.
FT   CHAIN      1712   1972       NS4B.
FT   CHAIN      1973   2419       NS5A.
FT   CHAIN      2420   3010       NS5B.
SQ   SEQUENCE   3010 AA;  327438 MW;  5F15AC675A0C8268 CRC64;

  Query Match             97.2%;  Score 1545;  DB 12;  Length 3010;
  Best Local Similarity   96.0%;  Pred. No. 6.1e-127;
  Matches  291;  Conservative    8;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||:|||||||||||||||||||| |||||||||:|||||||||:|||||||||| |
Db        904 AGLTKVPYFVRAQGLIRACMLVRKVAGGHYVQMAFVKLAALTGTYVYNHLTPLQDWAHTG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:||||| |||||||||||||||| ||||||||||||||||:||||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIISGLAVSARRGREILLGPADSFEGQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206
```

```
RESULT 4
Q9DTE6
ID   Q9DTE6      PRELIMINARY;      PRT;  3010 AA.
AC   Q9DTE6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCVT142;
RA   Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA   Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA   Mishiro S.;
RT   "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT   with hepatocellular carcinoma: the 'progression score' revisited.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AB049091; BAB18804.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW   Hydrolase; Nonstructural protein; Polyprotein;
KW   RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327042 MW;  3807DC6879684C95 CRC64;

  Query Match           97.2%;  Score 1544;  DB 12;  Length 3010;
  Best Local Similarity  95.7%;  Pred. No. 7.4e-127;
  Matches  290;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||:|||||||||||||||| |||||||||||||||||||||||||||:|||| |
Db        904 AGITRVPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLRDWAHTG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:||||| |||||||||||||||||||||||||||||||||: |||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIISGLPVSARRGREILLGPADSLEGQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              |||||||||||||||||||||||||||||||| |||||||||||||:|||:|||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHVVGIFRAAVCTRGVAKAVDFVPVEAMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 5
Q9DTE4
ID   Q9DTE4      PRELIMINARY;      PRT;  3010 AA.
AC   Q9DTE4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCVT150;
RA   Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA   Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA   Mishiro S.;
RT   "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT   with hepatocellular carcinoma: the 'progression score' revisited.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AB049093; BAB18806.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PQ0804; PQ0804.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
```

```
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327324 MW;  3DE6CF249BD1151C CRC64;

  Query Match           97.0%;  Score 1542;  DB 12;  Length 3010;
  Best Local Similarity  95.7%;  Pred. No. 1.1e-126;
  Matches  290;  Conservative    7;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        904 AGITKVPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAS 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:||||| |||||||||||||| |||||||||:||||||||: |||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCINGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              |||| ||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       1144 RRGDGRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 6
Q9DTD6
ID   Q9DTD6      PRELIMINARY;      PRT;  3010 AA.
AC   Q9DTD6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCVT221;
RA   Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA   Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA   Mishiro S.;
RT   "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT   with hepatocellular carcinoma: the 'progression score' revisited.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AB049101; BAB18814.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
```

```
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW   Hydrolase; Nonstructural protein; Polyprotein;
KW   RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327108 MW;  DE182D810EF78EE4 CRC64;

  Query Match           97.0%;  Score 1541;  DB 12;  Length 3010;
  Best Local Similarity  96.0%;  Pred. No. 1.4e-126;
  Matches  291;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              | | |||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db        904 AVIIKVPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |||||||||||||||||||||||||||||||||:||||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIISGLPVSARRGREILLGPADSFEGQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db       1024 RLLAPITAYSQQTRGLLGCIVTSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              |||||||||||||||||||||||||| |||||||:|||||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWPAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||:|||||||||||||||||||||||||| ||:|||||||||||||||||||||||||
Db       1144 RRGDTRGSLLSPRPVSYLKGSSGGPLLCPSGHIVGVFRAAVCTRGVAKAVDFIPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 7
P88803
ID   P88803      PRELIMINARY;      PRT;  3010 AA.
AC   P88803;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCV-1b;
RA   Enomoto N.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCV-1b;
RX   MEDLINE=95340824; PubMed=7542279;
RA   Enomoto N., Sakuma I., Asahina Y., Kurosaki M., Murakami T.,
RA   Yamamoto C., Izumi N., Marumo F., Sato C.;
RT   "Comparison of full-length sequences of interferon-sensitive and
RT   resistant hepatitis C virus 1b.";
RL   J. Clin. Invest. 96:224-230(1995).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; D50484; BAA09075.1; -.
DR   PIR; A61196; A61196.
DR   HSSP; P26663; 1NS3.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
FT   CHAIN         1    191       CORE PROTEIN.
FT   CHAIN       192    383       E1.
FT   CHAIN       384    809       E2.
FT   CHAIN       810   1026       NS2.
FT   CHAIN      1027   1657       NS3.
FT   CHAIN      1658   1711       NS4A.
FT   CHAIN      1712   1972       NS4B.
FT   CHAIN      1973   2419       NS5A.
FT   CHAIN      2420   3010       NS5B.
SQ   SEQUENCE   3010 AA;  327332 MW;  5F81505783FEFFB8 CRC64;

  Query Match           96.9%;  Score 1540;  DB 12;  Length 3010;
  Best Local Similarity  95.4%;  Pred. No. 1.7e-126;
  Matches  289;  Conservative    9;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||: :|||||||||||| |||||| ||||||||||:||||||||||:|||||||||| |
Db        904 AGLIRVPYFVRAQGLIRVCMLVRKVAGGHYVQMAFVKLAALTGTYVYNHLTPLQDWAHTG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |||||||||||||||||||||||||||||||||:||||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIISGLPVSARRGREILLGPADSFEGQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
```

```
                ||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        301 TMR 303
                |||
Db       1204 TMR 1206


RESULT 8
Q9J3H5
ID   Q9J3H5       PRELIMINARY;      PRT;  3010 AA.
AC   Q9J3H5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD17;
RA   Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT   "Characteristics of hepatitis C viral genome associated with disease
RT   progression.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF207758; AAF65948.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PQ0254; PQ0254.
DR   PIR; PS0329; PS0329.
DR   HSSP; P27958; 1HEI.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW   Hydrolase; Nonstructural protein; Polyprotein;
KW   RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  326801 MW;  9FEE3D1B93B7AA4B CRC64;

  Query Match           96.9%;  Score 1539;  DB 12;  Length 3010;
  Best Local Similarity  95.7%;  Pred. No. 2.1e-126;
  Matches  290;  Conservative    7;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||:|||||||||||||||| |||||||| |||||||||||||||||:|||| |
Db        904 AGITRVPYFVRAQGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHTG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |||||||||||||||||||||||||||||: |||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIISGLPVSARRGREILLGPADSLEGQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCINGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDLDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGLFRAAVCTRGVAKAVDFIPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 9
Q807P3
ID   Q807P3       PRELIMINARY;      PRT;  3010 AA.
AC   Q807P3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M1LE;
RX   MEDLINE=22047193; PubMed=12051758;
RA   Kishine H., Sugiyama K., Hijikata M., Kato N., Takahashi H., Noshi T.,
RA   Nio Y., Hosaka M., Miyanari Y., Shimotohno K.;
RT   "Subgenomic replicon derived from a cell line infected with the
RT   hepatitis C virus.";
RL   Biochem. Biophys. Res. Commun. 293:993-999(2002).
DR   EMBL; AB080299; BAC54896.1; -.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
```

```
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Polyprotein.
SQ   SEQUENCE   3010 AA;  327097 MW;  EE6418C7A723E686 CRC64;

  Query Match           96.9%;  Score 1539;  DB 12;  Length 3010;
  Best Local Similarity 96.0%;  Pred. No. 2.1e-126;
  Matches  291;  Conservative    5;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||:|||||||||||||||||||||||||||:||||||||||||| ||| ||| |
Db        904 AGITRVPYFVRAQGLIRACMLVRKAAGGHYVQMAFVKLAALTGTYVYDHLAPLQHWAHTG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |||||||||||||| ||||||||||||||||: |||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGPADSLEGQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCINGVCWTVFH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 RRGDGRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206
```

```
RESULT 10
Q9J3F9
ID   Q9J3F9      PRELIMINARY;      PRT;  3010 AA.
AC   Q9J3F9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD33;
RA   Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT   "Characteristics of hepatitis C viral genome associated with disease
RT   progression.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF207774; AAF65964.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PS0329; PS0329.
DR   HSSP; P27958; 1HEI.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
```

```
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327102 MW;  7162C9DB93E6E0C7 CRC64;

  Query Match             96.7%;  Score 1537;  DB 12;  Length 3010;
  Best Local Similarity   95.0%;  Pred. No. 3.1e-126;
  Matches  288;  Conservative   10;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||::|||||||||||||||||| ||||||||||||| |||||||:|||||||:||||||
Db        904 AGITRMPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLGALTGTYIYDHLTPLRDWAHAG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |||||||||||||||||||||:||||||| ||:||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIISGLPVSARRGKEILLGPADSLEGRGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              |||||||||||||||||||||||||||| ||||||:|||||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWPAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 11
O70815
ID   O70815      PRELIMINARY;      PRT;   361 AA.
AC   O70815;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Polyprotein (Fragment).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98321154; PubMed=9656998;
RA   Yamada K., Mori A., Seki M., Kimura J., Yuasa S., Matsuura Y.,
RA   Miyamura T.;
RT   "Critical point mutations for hepatitis C virus NS3 proteinase.";
RL   Virology 246:104-112(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Mori A., Yamada K., Kimura J., Koide T., Yuasa S., Yamada E.,
RA   Miyamura T.;
RT   "Enzymatic characterization of purified NS3 serine proteinase of
RT   hepatitis C virus expressed in Escherichia coli.";
RL   FEBS Lett. 378:37-42(1998).
DR   EMBL; AB013620; BAA28498.1; -.
DR   HSSP; P27958; 1HEI.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR004109; Peptidase_C29.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
FT   NON_TER       1      1
FT   NON_TER     361    361
SQ   SEQUENCE   361 AA;  38336 MW;  87DC310C76F4BCC3 CRC64;

  Query Match             96.5%;  Score 1534;  DB 12;  Length 361;
  Best Local Similarity   95.0%;  Pred. No. 3.5e-127;
  Matches  288;  Conservative    9;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||:|||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db          5 AGITRVPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHVS 64

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |||||||||||||||||||||||||||||:|||||:|| |||
Db         65 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIISGLPVSARRGREVLLGPADSFEHQGW 124

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|:|
Db        125 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWSVYH 184

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              |||||||||||:|||||||| ||||||||||||:||||||||||||||||||||||||||
Db        185 GAGSKTLAGPKGPVTQMYTNVDLDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 244

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db        245 RRGDSRGSLLSPRPISYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 304

Qy        301 TMR 303
              |||
Db        305 TMR 307


RESULT 12
Q9J3F4
ID   Q9J3F4      PRELIMINARY;      PRT;  3008 AA.
AC   Q9J3F4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
GN   MD34.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD34;
RA   Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT   "Characteristics of hepatitis C viral genome associated with disease
RT   progression.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF208024; AAF61205.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
```

```
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW   Hydrolase; Nonstructural protein; Polyprotein;
KW   RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3008 AA;  326834 MW;  99AE09E14C3109F4 CRC64;

  Query Match             96.5%;  Score 1534;  DB 12;  Length 3008;
  Best Local Similarity   95.7%;  Pred. No. 5.7e-126;
  Matches  290;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              | | |||||||||||||||||| || ||||||||||||||||||||||||||||||||||
Db        902 ASIIKVPYFVRAQGLIRACMLARKVAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAS 961

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:||||| |||||||||||||||| |||||||||:||||||||: |||||
Db        962 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGQGW 1021

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1022 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 1081

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db       1082 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1141

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       1142 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1201

Qy        301 TMR 303
              |||
Db       1202 TMR 1204


RESULT 13
Q9J3H3
```

```
ID   Q9J3H3      PRELIMINARY;      PRT;  3010 AA.
AC   Q9J3H3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD19;
RA   Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT   "Characteristics of hepatitis C viral genome associated with disease
RT   progression.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF207760; AAF65950.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
```

```
SQ   SEQUENCE   3010 AA;  327234 MW;  44C34677649CB8DD CRC64;

  Query Match             96.5%;  Score 1534;  DB 12;  Length 3010;
  Best Local Similarity   94.4%;  Pred. No. 5.7e-126;
  Matches  286;  Conservative   11;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||:|||||||||||||:| || ||||||||||||||||||||||||||||||||||||
Db        904 AGITRVPYFVRAQGLIRACLLARKVAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:||||| |||||||||||||||| |||||||||:|| |||||| |||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEIFLGPADNLEGQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||:||||||||||||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              |||:|||||||||||||||||||||||:||||||:||||||||||||||||||||||||
Db       1084 GAGAKTLAGPKGPITQMYTNVDQDLVGWQSPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||:|||||||||||||||||||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPISYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 14
Q9J3H2
ID   Q9J3H2      PRELIMINARY;      PRT;  3010 AA.
AC   Q9J3H2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD20;
RA   Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT   "Characteristics of hepatitis C viral genome associated with disease
RT   progression.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF207761; AAF65951.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1NS3.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  326763 MW;  1A48EE4BE51440D0 CRC64;

  Query Match             96.5%;  Score 1533;  DB 12;  Length 3010;
  Best Local Similarity   95.0%;  Pred. No. 7e-126;
  Matches  288;  Conservative    9;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              |||||||||||| |||||||||| |||||:|||||||||||||||||||:||||||
Db        904 AGITKVPYFVRAHGLIRACMLVRKVAGGHYLQMAFMKLAALTGTYVYDHLTPLRDWAHAG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |||||||||||||||||||||||||||||||||||: |||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIISGLPVSARRGREILLGPADSLEGQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              |||||||||:|||||||||||||||||||||||||||||||||||||||||||||||:||
Db       1024 RLLAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||| ||| |||:|||||||||||||||||:||||||||||||||||||||||||||||
Db       1084 GAGSKALAGQKGPVTQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 15
Q9J3I0
ID   Q9J3I0      PRELIMINARY;      PRT;  3010 AA.
AC   Q9J3I0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
```

```
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD12;
RA   Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT   "Characteristics of hepatitis C viral genome associated with disease
RT   progression.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF207753; AAF65943.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW   Hydrolase; Nonstructural protein; Polyprotein;
KW   RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  326692 MW;  074098DB305AF1A9 CRC64;

  Query Match           96.5%;  Score 1533;  DB 12;  Length 3010;
  Best Local Similarity   95.4%;  Pred. No. 7e-126;
  Matches  289;  Conservative    6;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||:|||||||||||||||||| |||||||||||||||||||||||: ||| |
Db        904 AGITRVPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPLRGWAHTG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |||||||||||||| |||||||||||||||: |||||
Db        964 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGPADSLEGQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQTFLATCVNGVCWTVFH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              |||| |||||||||||||||||||||||| |||||||||||||||||:|||||||
Db       1144 RRGDGRGSLLSPRPVSYLKGSSGGPLLCPSGHVVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


Search completed: May  6, 2004, 09:35:45
Job time : 29.339 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:25:16 ; Search time 11.8923 Seconds
(without alignments)
1315.364 Million cell updates/sec

Title: US-10-650-585-10
Perfect score: 1589
Sequence: 1 AGITKVPYFVRAQGLIRACM.....RGVAKAVDFIPVESMETTMR 303

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1530	96.3	2201	4	US-09-539-601-6	Sequence 6, Appli
2	1530	96.3	2201	4	US-09-539-601-15	Sequence 15, Appl
3	1530	96.3	3010	4	US-09-539-601-3	Sequence 3, Appli
4	1530	96.3	3010	4	US-09-539-601-21	Sequence 21, Appl
5	1530	96.3	3010	4	US-09-539-601-27	Sequence 27, Appl
6	1527	96.1	1692	3	US-09-263-933-4	Sequence 4, Appli
7	1527	96.1	1692	4	US-09-919-901-4	Sequence 4, Appli
8	1527	96.1	2307	3	US-09-263-933-2	Sequence 2, Appli
9	1527	96.1	2307	4	US-09-919-901-2	Sequence 2, Appli
10	1524	95.9	1692	3	US-09-263-933-11	Sequence 11, Appl
11	1524	95.9	1692	4	US-09-919-901-11	Sequence 11, Appl
12	1524	95.9	2307	3	US-09-263-933-9	Sequence 9, Appli
13	1524	95.9	2307	4	US-09-919-901-9	Sequence 9, Appli
14	1523	95.8	3010	4	US-09-539-601-33	Sequence 33, Appl
15	1515	95.3	1692	3	US-09-263-933-18	Sequence 18, Appl
16	1515	95.3	1692	4	US-09-919-901-18	Sequence 18, Appl
17	1515	95.3	2307	3	US-09-263-933-16	Sequence 16, Appl
18	1515	95.3	2307	4	US-09-919-901-16	Sequence 16, Appl
19	1504	94.7	3010	3	US-09-014-416-3	Sequence 3, Appli
20	1478	93.0	2013	1	US-08-324-977-12	Sequence 12, Appl
21	1478	93.0	2013	2	US-08-384-616-12	Sequence 12, Appl
22	1478	93.0	2013	2	US-08-904-686A-12	Sequence 12, Appl
23	1478	93.0	2013	3	US-09-315-850-12	Sequence 12, Appl
24	1478	93.0	2201	4	US-08-952-981A-2	Sequence 2, Appli
25	1478	93.0	2620	1	US-08-324-977-32	Sequence 32, Appl
26	1478	93.0	2620	2	US-08-384-616-32	Sequence 32, Appl
27	1478	93.0	2620	2	US-08-904-686A-32	Sequence 32, Appl
28	1478	93.0	2620	3	US-09-315-850-32	Sequence 32, Appl
29	1478	93.0	2621	1	US-08-324-977-36	Sequence 36, Appl
30	1478	93.0	2621	2	US-08-384-616-36	Sequence 36, Appl
31	1478	93.0	2621	2	US-08-904-686A-36	Sequence 36, Appl
32	1478	93.0	2621	3	US-09-315-850-36	Sequence 36, Appl
33	1478	93.0	3010	1	US-08-324-977-2	Sequence 2, Appli
34	1478	93.0	3010	1	US-08-324-977-14	Sequence 14, Appl
35	1478	93.0	3010	2	US-08-384-616-2	Sequence 2, Appli
36	1478	93.0	3010	2	US-08-384-616-14	Sequence 14, Appl
37	1478	93.0	3010	2	US-08-904-686A-2	Sequence 2, Appli
38	1478	93.0	3010	2	US-08-904-686A-14	Sequence 14, Appl
39	1478	93.0	3010	3	US-09-315-850-2	Sequence 2, Appli
40	1478	93.0	3010	3	US-09-315-850-14	Sequence 14, Appl
41	1405	88.4	3012	3	US-08-811-566-2	Sequence 2, Appli
42	1405	88.4	3012	4	US-09-034-756-2	Sequence 2, Appli
43	1402	88.2	1648	1	US-08-188-281B-12	Sequence 12, Appl
44	1402	88.2	1648	5	PCT-US94-07280-12	Sequence 12, Appl
45	1402	88.2	1648	5	PCT-US95-01087-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-539-601-6
; Sequence 6, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartenschlager, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2201
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-539-601-6

Query Match 96.3%; Score 1530; DB 4; Length 2201;
Best Local Similarity 95.0%; Pred. No. 2.5e-144;
Matches 288; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

```
Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              |||||||||||| |||||||||| ||||||||| ||||||||||||||||||:||||||
Db         95 AGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAG 154

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |:|||||||||||||| |||||||||| |||||: |||||
Db        155 LRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQGW 214

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||:|||||||||||||||||||||||||||||||:|
Db        215 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 274

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db        275 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 334

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        335 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 394

Qy        301 TMR 303
              |||
Db        395 TMR 397
```

RESULT 2

```
US-09-539-601-15
; Sequence 15, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
;  APPLICANT: Bartenschlager, Ralf FW
;  TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
;  FILE REFERENCE: all sequences
;  CURRENT APPLICATION NUMBER: US/09/539,601C
;  CURRENT FILING DATE: 2001-08-30
;  EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
;  EARLIER FILING DATE: 1999-04-03
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
;   LENGTH: 2201
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
US-09-539-601-15

  Query Match             96.3%;  Score 1530;  DB 4;  Length 2201;
  Best Local Similarity   95.0%;  Pred. No. 2.5e-144;
  Matches  288;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              |||||||||||| |||||||||| |||||||| ||||||||||||||||||:||||||
Db         95 AGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAG 154

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:||||| |:||||||||||||| |||||||||||| |||||: |||||
Db        155 LRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQGW 214

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||:|||||||||||||||||||||||||||||:||
Db        215 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 274

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||
Db        275 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 334

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        335 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 394

Qy        301 TMR 303
              |||
Db        395 TMR 397


RESULT 3
US-09-539-601-3
; Sequence 3, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
;  APPLICANT: Bartenschlager, Ralf FW
;  TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
;  FILE REFERENCE: all sequences
;  CURRENT APPLICATION NUMBER: US/09/539,601C
;  CURRENT FILING DATE: 2001-08-30
;  EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
;  EARLIER FILING DATE: 1999-04-03
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 3010
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
US-09-539-601-3

  Query Match             96.3%;  Score 1530;  DB 4;  Length 3010;
  Best Local Similarity   95.0%;  Pred. No. 3.9e-144;
  Matches  288;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              |||||||||||| |||||||||| |||||||| ||||||||||||||||||:||||||
Db        904 AGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:||||| |:||||||||||||| |||||||||||| |||||: |||||
Db        964 LRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||:|||||||||||||||||||||||||||||:||
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 4
US-09-539-601-21
; Sequence 21, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
;  APPLICANT: Bartenschlager, Ralf FW
;  TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
;  FILE REFERENCE: all sequences
;  CURRENT APPLICATION NUMBER: US/09/539,601C
;  CURRENT FILING DATE: 2001-08-30
;  EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
;  EARLIER FILING DATE: 1999-04-03
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
;   LENGTH: 3010
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
US-09-539-601-21

  Query Match             96.3%;  Score 1530;  DB 4;  Length 3010;
  Best Local Similarity   95.0%;  Pred. No. 3.9e-144;
  Matches  288;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              |||||||||||| |||||||||| |||||||| ||||||||||||||||||:||||||
Db        904 AGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:||||| |:||||||||||||| |||||||||||| |||||: |||||
Db        964 LRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||:|||||||||||||||||||||||||||||:||
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206
```

```
RESULT 5
US-09-539-601-27
; Sequence 27, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
;  APPLICANT: Bartenschlager, Ralf FW
;  TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
;  FILE REFERENCE: all sequences
;  CURRENT APPLICATION NUMBER: US/09/539,601C
;  CURRENT FILING DATE: 2001-08-30
;  EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
;  EARLIER FILING DATE: 1999-04-03
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
;   LENGTH: 3010
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
US-09-539-601-27

  Query Match            96.3%;  Score 1530;  DB 4;  Length 3010;
  Best Local Similarity  95.0%;  Pred. No. 3.9e-144;
  Matches  288;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              |||||||||||| |||||||||| ||||||||| ||||||||||||||||:||||||
Db        904 AGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |:||||||||||||| |||||||||| |||||: |||||
Db        964 LRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206


RESULT 6
US-09-263-933-4
; Sequence 4, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT: Jackson, Roberta L.
;  APPLICANT: Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/263,933
;  CURRENT FILING DATE: 1999-03-08
;  EARLIER APPLICATION NUMBER: 09/129,611
;  EARLIER FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 1692
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
US-09-263-933-4

  Query Match            96.1%;  Score 1527;  DB 3;  Length 1692;
  Best Local Similarity  94.4%;  Pred. No. 3.5e-144;
  Matches  286;  Conservative   11;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||:|||||||||||| ||||||| |||||||||||||| ||||||:|:|||||:||||||
Db        183 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWAHAG 242

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |||||||||||||| |||||||||:||||||||: ||:||
Db        243 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGRGW 302

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        303 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 362

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db        363 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 422

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        423 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 482

Qy        301 TMR 303
              |||
Db        483 TMR 485


RESULT 7
US-09-919-901-4
; Sequence 4, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT:  Jackson, Roberta L.
;  APPLICANT:  Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION:  OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/919,901
;  CURRENT FILING DATE:  2001-08-02
;  PRIOR APPLICATION NUMBER: 09/263,933
;  PRIOR FILING DATE: 1999-02-08
;  PRIOR APPLICATION NUMBER: 09/129,611
;  PRIOR FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 1692
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: :
US-09-919-901-4

  Query Match            96.1%;  Score 1527;  DB 4;  Length 1692;
  Best Local Similarity  94.4%;  Pred. No. 3.5e-144;
  Matches  286;  Conservative   11;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||:|||||||||||| ||||||| |||||||||||||| ||||||:|:|||||:||||||
Db        183 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWAHAG 242

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |||||||||||||| |||||||||:||||||||: ||:||
Db        243 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGRGW 302

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        303 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 362

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
```

```
Db        363 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 422

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        423 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 482

Qy        301 TMR 303
              |||
Db        483 TMR 485


RESULT 8
US-09-263-933-2
; Sequence 2, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT: Jackson, Roberta L.
;  APPLICANT: Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/263,933
;  CURRENT FILING DATE: 1999-03-08
;  EARLIER APPLICATION NUMBER: 09/129,611
;  EARLIER FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 2307
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
US-09-263-933-2

  Query Match             96.1%;  Score 1527;  DB 3;  Length 2307;
  Best Local Similarity   94.4%;  Pred. No. 5.4e-144;
  Matches  286;  Conservative   11;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||:|||||||||||| ||||||| |||||||||||||| ||||||:|:|||||:||||||
Db        275 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWAHAG 334

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |||||||||||||||| |||||||||:||||||||: ||:||
Db        335 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGRGW 394

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        395 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 454

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db        455 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 514

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        515 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 574

Qy        301 TMR 303
              |||
Db        575 TMR 577


RESULT 9
US-09-919-901-2
; Sequence 2, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT:  Jackson, Roberta L.
;  APPLICANT:  Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION:  OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/919,901
;  CURRENT FILING DATE:  2001-08-02
;  PRIOR APPLICATION NUMBER: 09/263,933
;  PRIOR FILING DATE: 1999-02-08
;  PRIOR APPLICATION NUMBER: 09/129,611
;  PRIOR FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 2307
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: :
US-09-919-901-2

  Query Match             96.1%;  Score 1527;  DB 4;  Length 2307;
  Best Local Similarity   94.4%;  Pred. No. 5.4e-144;
  Matches  286;  Conservative   11;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||:|||||||||||| ||||||| |||||||||||||| ||||||:|:|||||:||||||
Db        275 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWAHAG 334

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |||||||||||||||| |||||||||:||||||||: ||:||
Db        335 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGRGW 394

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        395 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 454

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db        455 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 514

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        515 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 574

Qy        301 TMR 303
              |||
Db        575 TMR 577


RESULT 10
US-09-263-933-11
; Sequence 11, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT: Jackson, Roberta L.
;  APPLICANT: Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/263,933
;  CURRENT FILING DATE: 1999-03-08
;  EARLIER APPLICATION NUMBER: 09/129,611
;  EARLIER FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 1692
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
US-09-263-933-11

  Query Match             95.9%;  Score 1524;  DB 3;  Length 1692;
  Best Local Similarity   94.1%;  Pred. No. 6.9e-144;
  Matches  285;  Conservative   12;  Mismatches    6;  Indels    0;  Gaps    0;
```

```
Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||:|||||||||||| ||||||| |||||||||||||| ||||||:|:|||||:||||||
Db        183 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWAHAG 242

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |||||||||||||| |||||||||:||||||||: ||:||
Db        243 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGRGW 302

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        303 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 362

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db        363 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 422

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||:|||||||||||||||||||||||||||||:|||||||
Db        423 RRGDSRGSLLSPRPVSYLKGSAGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 482

Qy        301 TMR 303
              |||
Db        483 TMR 485
```

RESULT 11
US-09-919-901-11
; Sequence 11, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: :
US-09-919-901-11

Query Match 95.9%; Score 1524; DB 4; Length 1692;
Best Local Similarity 94.1%; Pred. No. 6.9e-144;
Matches 285; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

```
Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||:|||||||||||| ||||||| |||||||||||||| ||||||:|:|||||:||||||
Db        183 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWAHAG 242

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |||||||||||||| |||||||||:||||||||: ||:||
Db        243 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGRGW 302

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        303 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 362

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db        363 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 422

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||:|||||||||||||||||||||||||||||:|||||||
Db        423 RRGDSRGSLLSPRPVSYLKGSAGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 482

Qy        301 TMR 303
              |||
Db        483 TMR 485
```

RESULT 12
US-09-263-933-9
; Sequence 9, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 2307
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-9

Query Match 95.9%; Score 1524; DB 3; Length 2307;
Best Local Similarity 94.1%; Pred. No. 1.1e-143;
Matches 285; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

```
Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||:|||||||||||| ||||||| |||||||||||||| ||||||:|:|||||:||||||
Db        275 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWAHAG 334

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |||||||||||||| |||||||||:||||||||: ||:||
Db        335 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGRGW 394

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        395 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 454

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db        455 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 514

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||:|||||||||||||||||||||||||||||:|||||||
Db        515 RRGDSRGSLLSPRPVSYLKGSAGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 574

Qy        301 TMR 303
              |||
Db        575 TMR 577
```

RESULT 13
US-09-919-901-9
; Sequence 9, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901

```
;  PRIOR APPLICATION NUMBER: 09/263,933
;  PRIOR FILING DATE: 1999-02-08
;  PRIOR APPLICATION NUMBER: 09/129,611
;  PRIOR FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 2307
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: :
US-09-919-901-9
```

Query Match 95.9%; Score 1524; DB 4; Length 2307;
Best Local Similarity 94.1%; Pred. No. 1.1e-143;
Matches 285; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

```
Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||:||||||||||| |||||| ||||||||||||| ||||||:|:|||||:||||||
Db        275 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWAHAG 334

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:||||| |||||||||||||| |||||||||:||||||||: ||:||
Db        335 LRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPADSLEGRGW 394

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        395 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 454

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db        455 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 514

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||:|||||||||||||||||||||||||||||:|||||||
Db        515 RRGDSRGSLLSPRPVSYLKGSAGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 574

Qy        301 TMR 303
              |||
Db        575 TMR 577
```

```
RESULT 14
US-09-539-601-33
; Sequence 33, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
;  APPLICANT: Bartenschlager, Ralf FW
;  TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
;  FILE REFERENCE: all sequences
;  CURRENT APPLICATION NUMBER: US/09/539,601C
;  CURRENT FILING DATE: 2001-08-30
;  EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
;  EARLIER FILING DATE: 1999-04-03
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
;   LENGTH: 3010
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
US-09-539-601-33
```

Query Match 95.8%; Score 1523; DB 4; Length 3010;
Best Local Similarity 94.7%; Pred. No. 2e-143;
Matches 287; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

```
Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              |||||||||||| |||||||||| |||||||||||||||||:|||||||
Db        904 AGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPLRDWAHAG 963

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||:||||| |:|||||||||||||| |||||||||||| |||||:|||||
Db        964 LRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPADSLEGQGW 1023

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||:|
Db       1024 RLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 1083

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db       1084 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 1143

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db       1144 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMGT 1203

Qy        301 TMR 303
              |||
Db       1204 TMR 1206
```

```
RESULT 15
US-09-263-933-18
; Sequence 18, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT: Jackson, Roberta L.
;  APPLICANT: Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/263,933
;  CURRENT FILING DATE: 1999-03-08
;  EARLIER APPLICATION NUMBER: 09/129,611
;  EARLIER FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
;   LENGTH: 1692
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
US-09-263-933-18
```

Query Match 95.3%; Score 1515; DB 3; Length 1692;
Best Local Similarity 93.7%; Pred. No. 5.6e-143;
Matches 284; Conservative 12; Mismatches 7; Indels 0; Gaps 0;

```
Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||:||||||||||| |||||| ||||||||||||| ||||||:|:|||||:||||||
Db        183 AGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPLRDWAHAG 242

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||:||||| |||||||||||| |||| |||||||||:||||||||: ||:||
Db        243 LRDLAVAVEPVVFSDMETKIITWGADTAAAGDIILGLPVSARRGKEILLGPADSLEGRGW 302

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        303 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVYH 362

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db        363 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVR 422

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||:|||||||||||||||||||||||||||||:|||||||
Db        423 RRGDSRGSLLSPRPVSYLKGSAGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFVPVESMET 482

Qy        301 TMR 303
              |||
Db        483 TMR 485
```

Search completed: May 6, 2004, 09:39:01
Job time : 12.8923 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:30:56 ; Search time 30.9958 Seconds
(without alignments)
2713.357 Million cell updates/sec

Title: US-10-650-585-10
Perfect score: 1589
Sequence: 1 AGITKVPYFVRAQGLIRACM.....RGVAKAVDFIPVESMETTMR 303

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1589	100.0	303	13	US-10-017-736-10	Sequence 10, Appl
2	1589	100.0	303	16	US-10-650-585-10	Sequence 10, Appl
3	1589	100.0	334	13	US-10-017-736-4	Sequence 4, Appli
4	1589	100.0	334	16	US-10-650-585-4	Sequence 4, Appli
5	1589	100.0	341	13	US-10-017-736-14	Sequence 14, Appl
6	1589	100.0	341	16	US-10-650-585-14	Sequence 14, Appl
7	1589	100.0	352	13	US-10-017-736-13	Sequence 13, Appl
8	1589	100.0	352	16	US-10-650-585-13	Sequence 13, Appl
9	1589	100.0	380	13	US-10-017-736-12	Sequence 12, Appl
10	1589	100.0	380	16	US-10-650-585-12	Sequence 12, Appl
11	1589	100.0	393	13	US-10-017-736-11	Sequence 11, Appl
12	1589	100.0	393	16	US-10-650-585-11	Sequence 11, Appl
13	1589	100.0	409	13	US-10-017-736-2	Sequence 2, Appli
14	1589	100.0	409	16	US-10-650-585-2	Sequence 2, Appli
15	1580	99.4	303	13	US-10-017-736-18	Sequence 18, Appl
16	1580	99.4	303	16	US-10-650-585-18	Sequence 18, Appl
17	1579	99.4	303	13	US-10-017-736-16	Sequence 16, Appl
18	1579	99.4	303	16	US-10-650-585-16	Sequence 16, Appl
19	1570	98.8	301	13	US-10-017-736-17	Sequence 17, Appl
20	1570	98.8	301	16	US-10-650-585-17	Sequence 17, Appl
21	1532	96.4	292	13	US-10-017-736-15	Sequence 15, Appl
22	1532	96.4	292	16	US-10-650-585-15	Sequence 15, Appl
23	1530	96.3	2201	13	US-10-029-907-3	Sequence 3, Appli
24	1530	96.3	2201	14	US-10-309-561-3	Sequence 3, Appli
25	1530	96.3	3010	12	US-10-467-000-1	Sequence 1, Appli
26	1527	96.1	1692	10	US-09-919-901-4	Sequence 4, Appli
27	1527	96.1	1692	14	US-10-191-966-4	Sequence 4, Appli
28	1527	96.1	2307	10	US-09-919-901-2	Sequence 2, Appli
29	1527	96.1	2307	14	US-10-191-966-2	Sequence 2, Appli
30	1524	95.9	1692	10	US-09-919-901-11	Sequence 11, Appl
31	1524	95.9	1692	14	US-10-191-966-11	Sequence 11, Appl
32	1524	95.9	2307	10	US-09-919-901-9	Sequence 9, Appli
33	1524	95.9	2307	14	US-10-191-966-9	Sequence 9, Appli
34	1515	95.3	1692	10	US-09-919-901-18	Sequence 18, Appl
35	1515	95.3	1692	14	US-10-191-966-18	Sequence 18, Appl
36	1515	95.3	2307	10	US-09-919-901-16	Sequence 16, Appl
37	1515	95.3	2307	14	US-10-191-966-16	Sequence 16, Appl
38	1478	93.0	2201	13	US-10-085-476-2	Sequence 2, Appli
39	1405	88.4	3011	9	US-09-742-659-4	Sequence 4, Appli
40	1405	88.4	3011	10	US-09-891-894-3	Sequence 3, Appli
41	1405	88.4	3011	14	US-10-184-150-3	Sequence 3, Appli
42	1405	88.4	3011	15	US-10-328-997-3	Sequence 3, Appli
43	1405	88.4	3012	9	US-09-238-076-2	Sequence 2, Appli
44	1405	88.4	3012	10	US-09-995-937-2	Sequence 2, Appli
45	1405	88.4	3012	10	US-09-917-563-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-10-017-736-10
; Sequence 10, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 303
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-10

Query Match 100.0%; Score 1589; DB 13; Length 303;
Best Local Similarity 100.0%; Pred. No. 1.5e-154;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300

Qy        301 TMR 303
              |||
Db        301 TMR 303


RESULT 2
US-10-650-585-10
; Sequence 10, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/650,585
;  CURRENT FILING DATE:  2003-08-28
;  PRIOR APPLICATION NUMBER: US/10/017,736A
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
;   LENGTH: 303
;   TYPE: PRT
;   ORGANISM: HCV
US-10-650-585-10

  Query Match             100.0%;  Score 1589;  DB 16;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-154;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300

Qy        301 TMR 303
              |||
Db        301 TMR 303


RESULT 3
US-10-017-736-4
; Sequence 4, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/017,736
;  CURRENT FILING DATE:  2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 334
;   TYPE: PRT
;   ORGANISM: HCV
US-10-017-736-4

  Query Match             100.0%;  Score 1589;  DB 13;  Length 334;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-154;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315

Qy        301 TMR 303
              |||
Db        316 TMR 318


RESULT 4
US-10-650-585-4
; Sequence 4, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/650,585
;  CURRENT FILING DATE:  2003-08-28
;  PRIOR APPLICATION NUMBER: US/10/017,736A
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 334
;   TYPE: PRT
;   ORGANISM: HCV
US-10-650-585-4

  Query Match             100.0%;  Score 1589;  DB 16;  Length 334;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-154;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         16 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 75

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         76 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 135

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
```

```
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 195

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 255

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 315

Qy        301 TMR 303
                  |||
Db        316 TMR 318


RESULT 5
US-10-017-736-14
; Sequence 14, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/017,736
;  CURRENT FILING DATE:  2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
;   LENGTH: 341
;   TYPE: PRT
;   ORGANISM: HCV
US-10-017-736-14

  Query Match             100.0%;  Score 1589;  DB 13;  Length 341;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-154;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         39 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 98

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 158

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 218

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 278

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 338

Qy        301 TMR 303
                  |||
Db        339 TMR 341


RESULT 6
US-10-650-585-14
; Sequence 14, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/650,585
;  CURRENT FILING DATE:  2003-08-28
;  PRIOR APPLICATION NUMBER: US/10/017,736A
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
;   LENGTH: 341
;   TYPE: PRT
;   ORGANISM: HCV
US-10-650-585-14

  Query Match             100.0%;  Score 1589;  DB 16;  Length 341;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-154;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         39 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 98

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 158

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 218

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 278

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 338

Qy        301 TMR 303
                  |||
Db        339 TMR 341


RESULT 7
US-10-017-736-13
; Sequence 13, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/017,736
;  CURRENT FILING DATE:  2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 352
;   TYPE: PRT
;   ORGANISM: HCV
US-10-017-736-13

  Query Match             100.0%;  Score 1589;  DB 13;  Length 352;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-154;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         50 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 109

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        110 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 169

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        170 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 229

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        230 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 289

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        290 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 349

Qy        301 TMR 303
              |||
Db        350 TMR 352


RESULT 8
US-10-650-585-13
; Sequence 13, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/650,585
;  CURRENT FILING DATE:  2003-08-28
;  PRIOR APPLICATION NUMBER: US/10/017,736A
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 352
;   TYPE: PRT
;   ORGANISM: HCV
US-10-650-585-13

  Query Match           100.0%;  Score 1589;  DB 16;  Length 352;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-154;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         50 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 109

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        110 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 169

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        170 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 229

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        230 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 289

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        290 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 349

Qy        301 TMR 303
              |||
Db        350 TMR 352


RESULT 9
US-10-017-736-12
; Sequence 12, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/017,736
;  CURRENT FILING DATE:  2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
;   LENGTH: 380
;   TYPE: PRT
;   ORGANISM: HCV
US-10-017-736-12

  Query Match           100.0%;  Score 1589;  DB 13;  Length 380;
  Best Local Similarity  100.0%;  Pred. No. 2e-154;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 137

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 197

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 257

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 317

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 377

Qy        301 TMR 303
              |||
Db        378 TMR 380


RESULT 10
US-10-650-585-12
; Sequence 12, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/650,585
;  CURRENT FILING DATE:  2003-08-28
;  PRIOR APPLICATION NUMBER: US/10/017,736A
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
;   LENGTH: 380
;   TYPE: PRT
;   ORGANISM: HCV
US-10-650-585-12

  Query Match           100.0%;  Score 1589;  DB 16;  Length 380;
  Best Local Similarity  100.0%;  Pred. No. 2e-154;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          78 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 137

Qy          61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         138 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 197

Qy         121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         198 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 257

Qy         181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         258 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 317

Qy         241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         318 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 377

Qy         301 TMR 303
                |||
Db         378 TMR 380


RESULT 11
US-10-017-736-11
; Sequence 11, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/017,736
;  CURRENT FILING DATE:  2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 393
;   TYPE: PRT
;   ORGANISM: HCV
US-10-017-736-11

  Query Match             100.0%;  Score 1589;  DB 13;  Length 393;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-154;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          91 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 150

Qy          61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         151 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 210

Qy         121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         211 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 270

Qy         181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         271 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 330

Qy         241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         331 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 390

Qy         301 TMR 303
                |||
Db         391 TMR 393


RESULT 12
US-10-650-585-11
; Sequence 11, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/650,585
;  CURRENT FILING DATE:  2003-08-28
;  PRIOR APPLICATION NUMBER: US/10/017,736A
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 393
;   TYPE: PRT
;   ORGANISM: HCV
US-10-650-585-11

  Query Match             100.0%;  Score 1589;  DB 16;  Length 393;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-154;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          91 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 150

Qy          61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         151 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 210

Qy         121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         211 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 270

Qy         181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         271 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 330

Qy         241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         331 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 390

Qy         301 TMR 303
                |||
Db         391 TMR 393


RESULT 13
US-10-017-736-2
; Sequence 2, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/017,736
;  CURRENT FILING DATE:  2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 409
;   TYPE: PRT
;   ORGANISM: HCV
US-10-017-736-2

  Query Match             100.0%;  Score 1589;  DB 13;  Length 409;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-154;
```

```
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         95 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 154

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        155 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 214

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        215 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 274

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        275 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 334

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        335 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 394

Qy        301 TMR 303
              |||
Db        395 TMR 397


RESULT 14
US-10-650-585-2
; Sequence 2, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/650,585
;  CURRENT FILING DATE:  2003-08-28
;  PRIOR APPLICATION NUMBER: US/10/017,736A
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 409
;   TYPE: PRT
;   ORGANISM: HCV
US-10-650-585-2

  Query Match            100.0%;  Score 1589;  DB 16;  Length 409;
  Best Local Similarity  100.0%;  Pred. No. 2.3e-154;
  Matches  303;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         95 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 154

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        155 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 214

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        215 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 274

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        275 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 334

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        335 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 394

Qy        301 TMR 303
              |||
Db        395 TMR 397


RESULT 15
US-10-017-736-18
; Sequence 18, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/017,736
;  CURRENT FILING DATE:  2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18
;   LENGTH: 303
;   TYPE: PRT
;   ORGANISM: HCV
US-10-017-736-18

  Query Match            99.4%;  Score 1580;  DB 13;  Length 303;
  Best Local Similarity  99.7%;  Pred. No. 1.2e-153;
  Matches  302;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAG 60

Qy         61 LRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQGW 120
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db         61 LRDLAVAVEPVIFSDMEVKIITWGADTAAAGDIISGLPVSARRGREILLGPADNFEGQGW 120

Qy        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFH 180

Qy        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVR 240

Qy        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300

Qy        301 TMR 303
              |||
Db        301 TMR 303


Search completed: May  6, 2004, 09:43:18
Job time : 30.9958 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:08:45 ; Search time 53.4939 Seconds
(without alignments)
2075.771 Million cell updates/sec

Title: US-10-650-585-11
Perfect score: 2053
Sequence: 1 MAASCGGAVFIGLALLTLSP.....RGVAKAVDFIPVESMETTMR 393

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2053	100.0	393	5	ABG32184	Abg32184 HCV prote
2	2053	100.0	409	5	ABG32181	Abg32181 HCV prote
3	1987	96.8	380	5	ABG32185	Abg32185 HCV prote
4	1967	95.8	3010	2	AAR82694	Aar82694 Partial H
5	1962	95.6	3010	2	AAR68622	Aar68622 HCV prote
6	1961	95.5	3010	2	AAR68864	Aar68864 Hepatitis
7	1951	95.0	2201	5	ABG30601	Abg30601 Hepatitis
8	1951	95.0	2201	5	ABG30591	Abg30591 Hepatitis
9	1951	95.0	2201	5	ABG30600	Abg30600 Hepatitis
10	1951	95.0	2201	5	ABG30581	Abg30581 Hepatitis
11	1951	95.0	2201	5	ABG30593	Abg30593 Hepatitis
12	1951	95.0	2201	5	ABG30582	Abg30582 Hepatitis
13	1951	95.0	2201	5	ABG30580	Abg30580 Hepatitis
14	1951	95.0	2201	5	ABG30587	Abg30587 Hepatitis
15	1951	95.0	2201	5	ABG30599	Abg30599 Hepatitis
16	1951	95.0	2201	5	ABG30594	Abg30594 Hepatitis
17	1951	95.0	2201	5	ABG30598	Abg30598 Hepatitis
18	1951	95.0	2201	5	ABG30595	Abg30595 Hepatitis
19	1951	95.0	3010	5	ABG32458	Abg32458 Hepatitis
20	1951	95.0	3010	5	ABG32459	Abg32459 Hepatitis
21	1951	95.0	3010	5	ABG32451	Abg32451 Hepatitis
22	1951	95.0	3010	5	ABG32455	Abg32455 Hepatitis
23	1951	95.0	3010	5	ABG32457	Abg32457 Hepatitis
24	1951	95.0	3010	5	ABG32460	Abg32460 Hepatitis
25	1951	95.0	3010	5	ABG32461	Abg32461 Hepatitis
26	1951	95.0	3010	5	ABG32454	Abg32454 Hepatitis
27	1951	95.0	3011	5	ABG32456	Abg32456 Hepatitis
28	1948	94.9	2201	5	ABG30586	Abg30586 Hepatitis
29	1948	94.9	2201	5	ABG30589	Abg30589 Hepatitis
30	1948	94.9	2201	5	ABG30583	Abg30583 Hepatitis
31	1948	94.9	2201	5	ABG30588	Abg30588 Hepatitis
32	1947	94.8	2201	5	ABG30590	Abg30590 Hepatitis
33	1946	94.8	2307	3	AAY70064	Aay70064 Recombina
34	1945	94.7	3010	5	ABG32452	Abg32452 Hepatitis
35	1944	94.7	2201	5	ABG30584	Abg30584 Hepatitis
36	1944	94.7	2201	5	ABG30602	Abg30602 Hepatitis
37	1944	94.7	3010	5	ABG32453	Abg32453 Hepatitis
38	1943	94.6	2307	3	AAY70065	Aay70065 Recombina
39	1943	94.6	3014	2	AAR54099	Aar54099 NANBHV E1
40	1940	94.5	2201	5	ABG30585	Abg30585 Hepatitis
41	1938.5	94.4	768	2	AAR40223	Aar40223 Recombina
42	1938	94.4	3014	2	AAR35207	Aar35207 Hepatitis
43	1937	94.3	3090	7	ADD67962	Add67962 EMCV inte
44	1934	94.2	2307	3	AAY70066	Aay70066 Recombina
45	1929	94.0	3010	5	AAE20477	Aae20477 HCV-S1 fu

ALIGNMENTS

RESULT 1
ABG32184
ID ABG32184 standard; protein; 393 AA.
XX
AC ABG32184;
XX
DT 05-NOV-2002 (first entry)
XX
DE HCV protease NS2/3 truncation mutant 815-1206.
XX
KW HCV; enzyme; protease; NS2/3 (815-1206); hepatitis C virus infection;
KW chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW chaotropic agent; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO200248375-A2.
XX
PD 20-JUN-2002.
XX
PF 13-DEC-2001; 2001WO-CA001796.
XX
PR 15-DEC-2000; 2000US-0256031P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
XX
DR WPI; 2002-599511/64.
XX
PT Novel polypeptide for screening inhibitors of non-structural proteases
PT useful as therapeutic agents against hepatitis C virus, comprises full
PT length non-structural protease, or its truncation.
XX
PS Claim 41; Page 59-60; 67pp; English.
XX
CC The invention relates to an isolated polypeptide consisting of a full-
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC residue amino acid 810 to 906, or having a minimal amino acid sequence
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC NS2/3 protease. Also included are (1) a composition (C) comprising an
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC its truncation or a mutated sequence, where the protease is in a solution
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide

CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC protease, involving isolating the protease in the presence of a
CC chaotropic agent, refolding the isolated protease by contacting it with a
CC reducing agent, and LDAO in the presence of reduced concentration of the
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC protease, involving diluting refolded inactive NS2/3 protease in a medium
CC containing an activation detergent to induce auto-cleavage of the NS2/3
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC protease, involving incubating the active NS2/3 protease produced by M2
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC cleavage products or their fragments, and measuring the presence or
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC active NS2/3 protease, involving carrying out M3 in the presence of, or
CC absence of the potential inhibitor, comparing the amount of uncleaved
CC NS2/3 protease, cleavage products or their fragments. The protease is
CC useful for detailed biochemical characterisation of the enzymes and in
CC the development of in vitro assays for screening novel inhibitors of
CC NS2/3 protease which are useful as therapeutic agents against HCV
CC infection (which causes chronic liver disease, cirrhosis and end-stage
CC liver disease. M1 is useful for high level production of protease. The
CC present sequence represents the NS2/3 truncation mutant 815-1206
CC (numbered relative to the full length NS2/3 protein)
XX
SQ Sequence 393 AA;

Query Match 100.0%; Score 2053; DB 5; Length 393;
Best Local Similarity 100.0%; Pred. No. 7.1e-191;
Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||||||||||||||
Db        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
```

RESULT 2
ABG32181
ID ABG32181 standard; protein; 409 AA.
XX
AC ABG32181;
XX
DT 05-NOV-2002 (first entry)
XX
DE HCV protease NS2/3 (810-1206).
XX
KW HCV; enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;
KW chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW chaotropic agent; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 398. .409
FT /note= "Streptavidin tag"
XX
PN WO200248375-A2.
XX
PD 20-JUN-2002.
XX
PF 13-DEC-2001; 2001WO-CA001796.
XX
PR 15-DEC-2000; 2000US-0256031P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
XX
DR WPI; 2002-599511/64.
DR N-PSDB; ABK90406.
XX
PT Novel polypeptide for screening inhibitors of non-structural proteases
PT useful as therapeutic agents against hepatitis C virus, comprises full
PT length non-structural protease, or its truncation.
XX
PS Claim 42; Fig 1B; 67pp; English.
XX
CC The invention relates to an isolated polypeptide consisting of a full-
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC residue amino acid 810 to 906, or having a minimal amino acid sequence
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC NS2/3 protease. Also included are (1) a composition (C) comprising an
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC its truncation or a mutated sequence, where the protease is in a solution
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC protease, involving isolating the protease in the presence of a
CC chaotropic agent, refolding the isolated protease by contacting it with a
CC reducing agent, and LDAO in the presence of reduced concentration of the
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC protease, involving diluting refolded inactive NS2/3 protease in a medium
CC containing an activation detergent to induce auto-cleavage of the NS2/3
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC protease, involving incubating the active NS2/3 protease produced by M2
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC cleavage products or their fragments, and measuring the presence or
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC active NS2/3 protease, involving carrying out M3 in the presence of, or
CC absence of the potential inhibitor, comparing the amount of uncleaved
CC NS2/3 protease, cleavage products or their fragments. The protease is
CC useful for detailed biochemical characterisation of the enzymes and in
CC the development of in vitro assays for screening novel inhibitors of
CC NS2/3 protease which are useful as therapeutic agents against HCV
CC infection (which causes chronic liver disease, cirrhosis and end-stage
CC liver disease. M1 is useful for high level production of protease. The
CC present sequence represents the NS2/3 (810-1206) protein, which has a C-
CC terminal streptavividin tag
XX
SQ Sequence 409 AA;

Query Match 100.0%; Score 2053; DB 5; Length 409;
Best Local Similarity 100.0%; Pred. No. 7.6e-191;
Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 64
```

```
Qy          61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          65 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 124

Qy         121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         125 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 184

Qy         181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         185 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 244

Qy         241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         245 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 304

Qy         301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         305 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 364

Qy         361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
               |||||||||||||||||||||||||||||||||
Db         365 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 397
```

RESULT 3
ABG32185
ID ABG32185 standard; protein; 380 AA.
XX
AC ABG32185;
XX
DT 05-NOV-2002 (first entry)
XX
DE HCV protease NS2/3 truncation mutant 827-1206.
XX
KW HCV; enzyme; protease; NS2/3 (827-1206); hepatitis C virus infection;
KW chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW chaotropic agent; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO200248375-A2.
XX
PD 20-JUN-2002.
XX
PF 13-DEC-2001; 2001WO-CA001796.
XX
PR 15-DEC-2000; 2000US-0256031P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
XX
DR WPI; 2002-599511/64.
XX
PT Novel polypeptide for screening inhibitors of non-structural proteases
PT useful as therapeutic agents against hepatitis C virus, comprises full
PT length non-structural protease, or its truncation.
XX
PS Claim 41; Page 60-61; 67pp; English.
XX
CC The invention relates to an isolated polypeptide consisting of a full-
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC residue amino acid 810 to 906, or having a minimal amino acid sequence
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC NS2/3 protease. Also included are (1) a composition (C) comprising an
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC its truncation or a mutated sequence, where the protease is in a solution
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC protease, involving isolating the protease in the presence of a
CC chaotropic agent, refolding the isolated protease by contacting it with a
CC reducing agent, and LDAO in the presence of reduced concentration of the
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC protease, involving diluting refolded inactive NS2/3 protease in a medium
CC containing an activation detergent to induce auto-cleavage of the NS2/3
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC protease, involving incubating the active NS2/3 protease produced by M2
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC cleavage products or their fragments, and measuring the presence or
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC active NS2/3 protease, involving carrying out M3 in the presence of, or
CC absence of the potential inhibitor, comparing the amount of uncleaved
CC NS2/3 protease, cleavage products or their fragments. The protease is
CC useful for detailed biochemical characterisation of the enzymes and in
CC the development of in vitro assays for screening novel inhibitors of
CC NS2/3 protease which are useful as therapeutic agents against HCV
CC infection (which causes chronic liver disease, cirrhosis and end-stage
CC liver disease. M1 is useful for high level production of protease. The
CC present sequence represents the NS2/3 truncation mutant 827-1206
CC (numbered relative to the full length NS2/3 protein)
XX
SQ Sequence 380 AA;

Query Match 96.8%; Score 1987; DB 5; Length 380;
Best Local Similarity 100.0%; Pred. No. 1.8e-184;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          14 ALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIF 73
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 ALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIF 60

Qy          74 DITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTG 133
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 DITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTG 120

Qy         134 TYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 193
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 TYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 180

Qy         194 GREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTAT 253
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTAT 240

Qy         254 QSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCG 313
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 QSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCG 300

Qy         314 SSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCT 373
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 SSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCT 360

Qy         374 RGVAKAVDFIPVESMETTMR 393
               ||||||||||||||||||||
Db         361 RGVAKAVDFIPVESMETTMR 380
```

RESULT 4
AAR82694
ID AAR82694 standard; protein; 3010 AA.
XX
AC AAR82694;
XX
DT 16-OCT-2003 (revised)
DT 14-NOV-1996 (first entry)
XX
DE Partial HCV non-structural polyprotein.
XX
KW proteinase; hepatitis C virus; screening; inhibitor; proteolytic;
KW identification; cleavage.
XX

OS Hepatitis C virus; Virus.
XX
FH Key Location/Qualifiers
FT Protein 898. .1233
FT /note= "partial proteinase; see AAR82692"
FT Protein 992. .1907
FT /note= "partial proteinase; see AAR82693"
XX
PN JP07184648-A.
XX
PD 25-JUL-1995.
XX
PF 05-FEB-1993; 93JP-00018854.
XX
PR 07-FEB-1992; 92JP-00022657.
PR 18-SEP-1992; 92JP-00249240.
PR 04-DEC-1992; 92JP-00325303.
XX
PA (KAEN/) KAENNO K.
PA (SUMQ) SUMITOMO METAL IND LTD.
PA (SOYA-) SOYAKU GIJUTSU KENKYUSHO KK.
XX
DR WPI; 1995-287962/38.
DR N-PSDB; AAT03960.
XX
PT An HCV proteinase active substance - which has activity as an anti-HCV
PT agent and can be used to screen for proteinase inhibitors.
XX
PS Disclosure; Page 39-48; 52pp; Japanese.
XX
CC The present sequence is a partial Hepatitis C Virus (HCV) polyprotein
CC from the non-structral region. Partial proteinase sequences (AAR82692-
CC 93) are contained within this sequence. The proteinases can be used as
CC anti-HCV agents. They can also be used to screen cpds. for their ability
CC to inhibit their proteolytic activity. In this way proteinase inhibitors
CC can be identified. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ Sequence 3010 AA;

Query Match 95.8%; Score 1967; DB 2; Length 3010;
Best Local Similarity 93.9%; Pred. No. 3.1e-181;
Matches 369; Conservative 13; Mismatches 11; Indels 0; Gaps 0;

```
Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| |||||||||| |||||||||| ||| |||||||:||||||||||||
Db        814 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFITRAEAHLQVWVPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              |||||||||||||||||||||| ||||||||||:|||||||||||||||||||| |||||
Db        874 ILLTCAVHPELIFDITKLLLAILGPLMVLQAGITRVPYFVRAQGLIRACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              ||||||||||||||||||||:||||||||||||||||:||||| |:||||||||||
Db        934 VQMAFMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKLITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              ||||||||||||:||||||:|  ||||||||||||||||||||||||||||||||||||
Db        994 GDIISGLPVSARRGKEILLGPADSFGEQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||:|||||:|||||||||||||||||:|||||||||||||||||||||||||||||| |
Db       1054 QVDGEVQVLSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWPA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||||||||||||||||||||:|||||||||||||:|||||||||||||||||||||
Db       1114 PPGARSMTPCTCGSSDLYLVTRHADVVPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              || |||||||||||||||||||||||||||||
Db       1174 GHVVGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206
```

RESULT 5

AAR68622
ID AAR68622 standard; protein; 3010 AA.
XX
AC AAR68622;
XX
DT 16-OCT-2003 (revised)
DT 16-OCT-1995 (first entry)
XX
DE HCV protein cleavable with new serine proteinase.
XX
KW proteinase; serine; cleavage; hepatitis C virus; HCV.
XX
OS Hepatitis C virus; Virus.
XX
FH Key Location/Qualifiers
FT Cleavage-site 2419. .2420
FT /note= "Serine protease cleavage site"
XX
PN JP06315377-A.
XX
PD 15-NOV-1994.
XX
PF 06-MAY-1993; 93JP-00105666.
XX
PR 06-MAY-1993; 93JP-00105666.
XX
PA (KAEN/) KAENNO K.
PA (SUMQ) SUMITOMO METAL IND LTD.
PA (SOYA-) SOYAKU GIJUTSU KENKYUSHO KK.
XX
DR WPI; 1995-032330/05.
DR N-PSDB; AAQ80498.
XX
PT New HCV-originated proteinase active substance - used for site-specific
PT cleavage by an intermolecular reaction and the purification thereof.
XX
PS Disclosure; Page 10-19; 23pp; Japanese.
XX
CC This protein from HCV (hepatitis C virus) (encoded by AAQ80498) is
CC cleaved between amino acids 2419 and 2420, by a new serine protease,
CC contg. the sequence of AAR68621. The proteinase is purified as a fused
CC product with the dihydrofolate reductase protein by using a methotrexate
CC column. It can be used for the development of an inhibitor for HCV
CC proteinase. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ Sequence 3010 AA;

Query Match 95.6%; Score 1962; DB 2; Length 3010;
Best Local Similarity 93.6%; Pred. No. 9.4e-181;
Matches 368; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

```
Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| |||||||||| |||||||||| ||| |||||||:||||||||||||
Db        814 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFITRAEAHLQVWVPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              |||||||||||||||||||| | |||||||||||:|||||||||||||||||| |||||
Db        874 ILLTCAVHPELIFDITKLLLPILGPLMVLQAGITRVPYFVRAQGLIRACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              ||||||||||||||||||||:||||||||||||||||:||||| |:||||||||||
Db        934 VQMAFMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKLITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              ||||||||||||:||||||:|  ||||||||||||||||||||||||||||||||||||
Db        994 GDIISGLPVSARRGKEILLGPADSFGEQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||:||||:|||||||||||||||||:|||||||||||||||||||||||||||||| |
Db       1054 QVDGEVQVLSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWPA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||||||||||||||||||||:|||||||||||||:|||||||||||||||||||||
```

```
Db       1114 PPGARSMTPCTCGSSDLYLVTRHADVVPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              || ||||||||||||||||||||||||||||||||
Db       1174 GHVVGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206


RESULT 6
AAR68864
ID   AAR68864 standard; protein; 3010 AA.
XX
AC   AAR68864;
XX
DT   06-DEC-1995  (first entry)
XX
DE   Hepatitis C virus RNA helicase.
XX
KW   Hepatitis C virus; HCV; non-A non-B; helicase gene; RNA helicase;
KW   baculovirus; recombinant production.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   Region          196. .198
FT                   /label= N-linked glycosylation site
FT   Region          209. .211
FT                   /label= N-linked glycosylation site
FT   Region          234. .236
FT                   /label= N-linked glycosylation site
FT   Region          250. .252
FT                   /label= N-linked glycosylation site
FT   Region          305. .307
FT                   /label= N-linked glycosylation site
FT   Region          325. .327
FT                   /label= N-linked glycosylation site
FT   Region          417. .419
FT                   /label= N-linked glycosylation site
FT   Region          423. .425
FT                   /label= N-linked glycosylation site
FT   Region          430. .432
FT                   /label= N-linked glycosylation site
FT   Region          448. .450
FT                   /label= N-linked glycosylation site
FT   Region          532. .534
FT                   /label= N-linked glycosylation site
FT   Region          556. .558
FT                   /label= N-linked glycosylation site
FT   Region          576. .578
FT                   /label= N-linked glycosylation site
FT   Region          623. .625
FT                   /label= N-linked glycosylation site
FT   Region          645. .647
FT                   /label= N-linked glycosylation site
FT   Region          1213. .1215
FT                   /label= N-linked glycosylation site
FT   Region          1255. .1257
FT                   /label= N-linked glycosylation site
FT   Region          2041. .2043
FT                   /label= N-linked glycosylation site
FT   Region          2077. .2079
FT                   /label= N-linked glycosylation site
FT   Region          2240. .2242
FT                   /label= N-linked glycosylation site
FT   Region          2788. .2790
FT                   /label= N-linked glycosylation site
XX
PN   JP06319583-A.
XX
PD   22-NOV-1994.
XX
PF   18-SEP-1992;   92JP-00249241.
XX
PR   18-SEP-1992;   92JP-00249241.
XX
PA   (SOYA-) SOYAKU GIJUTSU KENKYUSHO KK.
XX
DR   WPI; 1995-040330/06.
DR   N-PSDB; AAQ81559.
XX
PT   of hepatitis C virus helicase gene in baculovirus - useful for large
PT   scale prodn. of RNA helicase.
XX
PS   Claim 1; Fig 1-4; 9pp; Japanese.
XX
CC   AAQ81559 encodes AAR68864 hepatitis C virus (HCV) RNA helicase. The DNA
CC   was used in the construction of an expression vector, which was used to
CC   transform a baculovirus host. The transformed baculovirus could then be
CC   used for the recombinant prodn. of HCV RNA helicase
XX
SQ   Sequence 3010 AA;

  Query Match             95.5%;  Score 1961;  DB 2;  Length 3010;
  Best Local Similarity   93.6%;  Pred. No. 1.2e-180;
  Matches  368;  Conservative   13;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||:|| |||||||||| |||||||||| ||| |||||||:||||||||||||
Db        814 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFITRAEAHLQVWVPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              |||||||||||||||||||| ||||||||||:|||||||||||||||| |||||
Db        874 ILLTCAVHPELIFDITKLLLAILGPLMVLQAGITRVPYFVRAQGLIRACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||||||||||||||||:||||||||||||||||:||||| |:||||||||||
Db        934 VQMAFMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKLITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||||||||||||:|||||||:|  ||||||||||||||||||||||||||||||||||
Db        994 GDIISGLPVSARRGKEILLGPADSFGEQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||:|||||:||||||||||||||||:|||||||||| |||||||||||||||||| |
Db       1054 QVDGEVQVLSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKAPITQMYTNVDQDLVGWPA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              |||||||||||||||||||||||||||:|||||||||||||||:|||||||||||||||
Db       1114 PPGARSMTPCTCGSSDLYLVTRHADVVPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              || ||||||||||||||||||||||||||||||||
Db       1174 GHVVGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206


RESULT 7
ABG30601
ID   ABG30601 standard; protein; 2201 AA.
XX
AC   ABG30601;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #10.
XX
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 882
FT                   /label= Arg, Lys
FT   Misc-difference 2183
FT                   /note= "Wild type Met substituted by Thr"
XX
```

PN WO200252015-A2.
XX
PD 04-JUL-2002.
XX
PF 20-DEC-2001; 2001WO-CA001843.
XX
PR 22-DEC-2000; 2000US-0257857P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Kukolj G, Pause A;
XX
DR WPI; 2002-575382/61.
XX
PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT possess enhanced transduction or replication efficiency, useful for
PT evaluating potential inhibitors of HCV replication.
XX
PS Claim 3; Page; 140pp; English.
XX
CC The invention describes a self-replicating hepatitis C virus (HCV)
CC polynucleotide molecule comprising a 5'-non translated region (NTR),
CC where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC potential inhibitors of HCV replication. The HCV RNA molecule is also
CC useful for efficiently establishing cell culture replication. The self-
CC replicating polynucleotide molecule contains a 5'-NTR, where G at
CC position 1 is substituted for A, and therefore provides an alternative to
CC existing systems comprising a self-replicating HCV RNA molecule that, in
CC conjunction with mutations in the HCV non-structural region, such as the
CC G(2042)C/R mutations, transduces and/or replicates with greater
CC efficiency. This amino acid sequence represents a mutant of the hepatitis
CC C virus replicon APGK12 and contains the viral protease NS2/3, protease
CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
CC This sequence does not appear in the specification but has been created
CC from the wild type sequence shown in ABG30580 using information given in
CC the claims of the invention
XX
SQ Sequence 2201 AA;

Query Match 95.0%; Score 1951; DB 5; Length 2201;
Best Local Similarity 93.1%; Pred. No. 7.1e-180;
Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

```
Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||:|| |||||:||: |||||||||| ||| ||||||||||||||||||:
Db          5 MAASCGGAVFVGLILLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAV 64

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||:|||||| |||:|||| ||||||||||||||||| |||||||||| |||||
Db         65 ILLTCAIHPELIFTITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHY 124

Qy        121 VQMAFMKLAAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| ||||||||||||||||:||||||||||||||||:||||| |:||||||||||
Db        125 VQMALMKLAAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAAC 184

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||| |||||: ||||||||||||||||||||||||||||||:|
Db        185 GDIILGLPVSARRGREIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRN 244

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db        245 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 304

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 364

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||:||||||||||
Db        365 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 397
```

RESULT 8
ABG30591
ID ABG30591 standard; protein; 2201 AA.
XX
AC ABG30591;
XX
DT 21-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #3.
XX
KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 751
FT /note= "Wild type Ser substituted by Gly"
FT Misc-difference 882
FT /label= Arg, Lys
XX
PN WO200252015-A2.
XX
PD 04-JUL-2002.
XX
PF 20-DEC-2001; 2001WO-CA001843.
XX
PR 22-DEC-2000; 2000US-0257857P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Kukolj G, Pause A;
XX
DR WPI; 2002-575382/61.
XX
PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT possess enhanced transduction or replication efficiency, useful for
PT evaluating potential inhibitors of HCV replication.
XX
PS Claim 3; Page; 140pp; English.
XX
CC The invention describes a self-replicating hepatitis C virus (HCV)
CC polynucleotide molecule comprising a 5'-non translated region (NTR),
CC where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC potential inhibitors of HCV replication. The HCV RNA molecule is also
CC useful for efficiently establishing cell culture replication. The self-
CC replicating polynucleotide molecule contains a 5'-NTR, where G at
CC position 1 is substituted for A, and therefore provides an alternative to
CC existing systems comprising a self-replicating HCV RNA molecule that, in
CC conjunction with mutations in the HCV non-structural region, such as the
CC G(2042)C/R mutations, transduces and/or replicates with greater
CC efficiency. This amino acid sequence represents a mutant of the hepatitis
CC C virus replicon APGK12 and contains the viral protease NS2/3, protease
CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
CC This sequence does not appear in the specification but has been created
CC from the wild type sequence shown in ABG30580 using information given in
CC the claims of the invention
XX
SQ Sequence 2201 AA;

Query Match 95.0%; Score 1951; DB 5; Length 2201;
Best Local Similarity 93.1%; Pred. No. 7.1e-180;
Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

```
Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||:|| |||||:||: |||||||||| ||| ||||||||||||||||||:
Db          5 MAASCGGAVFVGLILLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAV 64

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
```

```
                ||||||:|||||| |||:|||| |||||||||||||||||| |||||||||| |||||
Db         65 ILLTCAIHPELIFTITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHY 124

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
                |||| |||||||||||||||:|||||||||||||||:||||| |:||||||||||
Db        125 VQMALMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAAC 184

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
                |||| |||||||||||| |||||: |||||||||||||||||||||||||||||||:|
Db        185 GDIILGLPVSARRGREIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRN 244

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
                ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db        245 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 304

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
                |||||||:|||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 364

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
                |||||||||||||||||||||:|||||||||
Db        365 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 397
```

RESULT 9
ABG30600
ID ABG30600 standard; protein; 2201 AA.
XX
AC ABG30600;
XX
DT 21-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #9.
XX
KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key   Location/Qualifiers
FT Misc-difference 882
FT   /label= Arg, Lys
FT Misc-difference 1357
FT   /note= "Wild type Pro substituted by Leu"
XX
PN WO200252015-A2.
XX
PD 04-JUL-2002.
XX
PF 20-DEC-2001; 2001WO-CA001843.
XX
PR 22-DEC-2000; 2000US-0257857P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Kukolj G, Pause A;
XX
DR WPI; 2002-575382/61.
XX
PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT possess enhanced transduction or replication efficiency, useful for
PT evaluating potential inhibitors of HCV replication.
XX
PS Claim 3; Page; 140pp; English.
XX
CC The invention describes a self-replicating hepatitis C virus (HCV)
CC polynucleotide molecule comprising a 5'-non translated region (NTR),
CC where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC potential inhibitors of HCV replication. The HCV RNA molecule is also
CC useful for efficiently establishing cell culture replication. The self-
CC replicating polynucleotide molecule contains a 5'-NTR, where G at
CC position 1 is substituted for A, and therefore provides an alternative to
CC existing systems comprising a self-replicating HCV RNA molecule that, in
CC conjunction with mutations in the HCV non-structural region, such as the
CC G(2042)C/R mutations, transduces and/or replicates with greater
CC efficiency. This amino acid sequence represents a mutant of the hepatitis
CC C virus replicon APGK12 and contains the viral protease NS2/3, protease
CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
CC This sequence does not appear in the specification but has been created
CC from the wild type sequence shown in ABG30580 using information given in
CC the claims of the invention
XX
SQ Sequence 2201 AA;

```
  Query Match             95.0%;  Score 1951;  DB 5;  Length 2201;
  Best Local Similarity   93.1%;  Pred. No. 7.1e-180;
  Matches  366;  Conservative   14;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||||:|| |||||:||: |||||||||| ||| |||||||||||||||||:
Db          5 MAASCGGAVFVGLILLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAV 64

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||:|||||| |||:|||| |||||||||||||||| |||||||||| |||||
Db         65 ILLTCAIHPELIFTITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHY 124

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
                |||| |||||||||||||||:|||||||||||||||:||||| |:||||||||||
Db        125 VQMALMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAAC 184

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
                |||| |||||||||||| |||||: |||||||||||||||||||||||||||||||:|
Db        185 GDIILGLPVSARRGREIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRN 244

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
                ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db        245 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 304

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
                |||||||:|||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 364

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
                |||||||||||||||||||||:|||||||||
Db        365 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 397
```

RESULT 10
ABG30581
ID ABG30581 standard; protein; 2201 AA.
XX
AC ABG30581;
XX
DT 21-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #1.
XX
KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW cell culture replication; NS2/3; NS3/4; NS3; NS5B.
XX
OS Hepatitis C virus.
XX
PN WO200252015-A2.
XX
PD 04-JUL-2002.
XX
PF 20-DEC-2001; 2001WO-CA001843.
XX
PR 22-DEC-2000; 2000US-0257857P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Kukolj G, Pause A;

```
XX
DR   WPI; 2002-575382/61.
DR   N-PSDB; ABK88573.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
XX
PS   Disclosure; Page 49-58; 140pp; English.
XX
CC   The invention describes a self-replicating hepatitis C virus (HCV)
CC   polynucleotide molecule comprising a 5'-non translated region (NTR),
CC   where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC   region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC   replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC   potential inhibitors of HCV replication. The HCV RNA molecule is also
CC   useful for efficiently establishing cell culture replication. The self-
CC   replicating polynucleotide molecule contains a 5'-NTR, where G at
CC   position 1 is substituted for A, and therefore provides an alternative to
CC   existing systems comprising a self-replicating HCV RNA molecule that, in
CC   conjunction with mutations in the HCV non-structural region, such as the
CC   G(2042)C/R mutations, transduces and/or replicates with greater
CC   efficiency. This amino acid sequence is encoded by the hepatitis C virus
CC   replicon APGK12 and contains the viral protease NS2/3, protease complex
CC   NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B
XX
SQ   Sequence 2201 AA;

  Query Match             95.0%;  Score 1951;  DB 5;  Length 2201;
  Best Local Similarity   93.1%;  Pred. No. 7.1e-180;
  Matches  366;  Conservative   14;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| ||||||:||: |||||||||| ||| ||||||||||||||||||:
Db          5 MAASCGGAVFVGLILLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAV 64

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||:|||||| |||:|||| |||||||||||||||||| |||||||||| |||||
Db         65 ILLTCAIHPELIFTITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHY 124

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| ||||||||||||||||:|||||||||||||||||:||||| |:||||||||||
Db        125 VQMALMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAAC 184

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||| |||||: |||||||||||||||||||||||||||||||:|
Db        185 GDIILGLPVSARRGREIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRN 244

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db        245 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 304

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 364

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
Db        365 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 397


RESULT 11
ABG30593
ID   ABG30593 standard; protein; 2201 AA.
XX
AC   ABG30593;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #4.
XX
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 882
FT                   /label= Arg, Lys
FT   Misc-difference 892
FT                   /note= "Wild type Leu substituted by Phe"
XX
PN   WO200252015-A2.
XX
PD   04-JUL-2002.
XX
PF   20-DEC-2001; 2001WO-CA001843.
XX
PR   22-DEC-2000; 2000US-0257857P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Kukolj G,  Pause A;
XX
DR   WPI; 2002-575382/61.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
XX
PS   Claim 3; Page; 140pp; English.
XX
CC   The invention describes a self-replicating hepatitis C virus (HCV)
CC   polynucleotide molecule comprising a 5'-non translated region (NTR),
CC   where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC   region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC   replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC   potential inhibitors of HCV replication. The HCV RNA molecule is also
CC   useful for efficiently establishing cell culture replication. The self-
CC   replicating polynucleotide molecule contains a 5'-NTR, where G at
CC   position 1 is substituted for A, and therefore provides an alternative to
CC   existing systems comprising a self-replicating HCV RNA molecule that, in
CC   conjunction with mutations in the HCV non-structural region, such as the
CC   G(2042)C/R mutations, transduces and/or replicates with greater
CC   efficiency. This amino acid sequence represents a mutant of the hepatitis
CC   C virus replicon APGK12 and contains the viral protease NS2/3, protease
CC   complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
CC   This sequence does not appear in the specification but has been created
CC   from the wild type sequence shown in ABG30580 using information given in
CC   the claims of the invention
XX
SQ   Sequence 2201 AA;

  Query Match             95.0%;  Score 1951;  DB 5;  Length 2201;
  Best Local Similarity   93.1%;  Pred. No. 7.1e-180;
  Matches  366;  Conservative   14;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| ||||||:||: |||||||||| ||| ||||||||||||||||||:
Db          5 MAASCGGAVFVGLILLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAV 64

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||:|||||| |||:|||| |||||||||||||||||| |||||||||| |||||
Db         65 ILLTCAIHPELIFTITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHY 124

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| ||||||||||||||||:|||||||||||||||||:||||| |:||||||||||
Db        125 VQMALMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAAC 184

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||| |||||: |||||||||||||||||||||||||||||||:|
Db        185 GDIILGLPVSARRGREIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRN 244

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
```

```
Db        245 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 304

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 364

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
Db        365 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 397


RESULT 12
ABG30582
ID   ABG30582 standard; protein; 2201 AA.
XX
AC   ABG30582;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #2.
XX
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 882
FT                   /note= "Wild type Lys substituted by Lys or Arg"
FT   Misc-difference 1233
FT                   /note= "Wild type Gly substituted by Cys"
XX
PN   WO200252015-A2.
XX
PD   04-JUL-2002.
XX
PF   20-DEC-2001; 2001WO-CA001843.
XX
PR   22-DEC-2000; 2000US-0257857P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Kukolj G,  Pause A;
XX
DR   WPI; 2002-575382/61.
DR   N-PSDB; ABK88574.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
XX
PS   Disclosure; Page 59-69; 140pp; English.
XX
CC   The invention describes a self-replicating hepatitis C virus (HCV)
CC   polynucleotide molecule comprising a 5'-non translated region (NTR),
CC   where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC   region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC   replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC   potential inhibitors of HCV replication. The HCV RNA molecule is also
CC   useful for efficiently establishing cell culture replication. The self-
CC   replicating polynucleotide molecule contains a 5'-NTR, where G at
CC   position 1 is substituted for A, and therefore provides an alternative to
CC   existing systems comprising a self-replicating HCV RNA molecule that, in
CC   conjunction with mutations in the HCV non-structural region, such as the
CC   G(2042)C/R mutations, transduces and/or replicates with greater
CC   efficiency. This amino acid sequence is encoded by the hepatitis C virus
CC   replicon APGK12 and contains the viral protease NS2/3, protease complex
CC   NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this
CC   sequence has been created from replicon APGK12 shown in ABG30581
XX
SQ   Sequence 2201 AA;

  Query Match             95.0%;  Score 1951;  DB 5;  Length 2201;
  Best Local Similarity   93.1%;  Pred. No. 7.1e-180;
  Matches  366;  Conservative   14;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| ||||||:||: |||||||||| ||| |||||||||||||||||:
Db          5 MAASCGGAVFVGLILLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAV 64

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||:|||||| |||:|||| |||||||||||||||||| |||||||||| |||||
Db         65 ILLTCAIHPELIFTITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHY 124

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| |||||||||||||||:|||||||||||||||:||||| |:||||||||||
Db        125 VQMALMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAAC 184

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||| |||||: |||||||||||||||||||||||||||||||:|
Db        185 GDIILGLPVSARRGREIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRN 244

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
Db        245 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 304

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 364

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
Db        365 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 397


RESULT 13
ABG30580
ID   ABG30580 standard; protein; 2201 AA.
XX
AC   ABG30580;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #9.
XX
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; NS2/3; NS3/4; NS3; NS5B.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 882
FT                   /note= "Encoded by ARG"
XX
PN   WO200252015-A2.
XX
PD   04-JUL-2002.
XX
PF   20-DEC-2001; 2001WO-CA001843.
XX
PR   22-DEC-2000; 2000US-0257857P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Kukolj G,  Pause A;
XX
DR   WPI; 2002-575382/61.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
XX
PS   Disclosure; Page 69-74; 140pp; English.
XX
CC   The invention describes a self-replicating hepatitis C virus (HCV)
CC   polynucleotide molecule comprising a 5'-non translated region (NTR),
```

CC where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC potential inhibitors of HCV replication. The HCV RNA molecule is also
CC useful for efficiently establishing cell culture replication. The self-
CC replicating polynucleotide molecule contains a 5'-NTR, where G at
CC position 1 is substituted for A, and therefore provides an alternative to
CC existing systems comprising a self-replicating HCV RNA molecule that, in
CC conjunction with mutations in the HCV non-structural region, such as the
CC G(2042)C/R mutations, transduces and/or replicates with greater
CC efficiency. This amino acid sequence is encoded by the hepatitis C virus
CC replicon APGK12 and contains the viral protease NS2/3, protease complex
CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B
XX
SQ Sequence 2201 AA;

Query Match 95.0%; Score 1951; DB 5; Length 2201;
Best Local Similarity 93.1%; Pred. No. 7.1e-180;
Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

```
Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| ||||||:||: |||||||||| ||| ||||||||||||||||||:
Db          5 MAASCGGAVFVGLILLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAV 64

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||:|||||| |||:|||| |||||||||||||||||| |||||||||| |||||
Db         65 ILLTCAIHPELIFTITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHY 124

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| |||||||||||||||:||||||||||||||||:||||| |:||||||||||
Db        125 VQMALMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAAC 184

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||| |||||: |||||||||||||||||||||||||||||||||:|
Db        185 GDIILGLPVSARRGREIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRN 244

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              |||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db        245 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 304

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 364

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||||:||||||||||
Db        365 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 397
```

RESULT 14
ABG30587
ID ABG30587 standard; protein; 2201 AA.
XX
AC ABG30587;
XX
DT 21-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #7.
XX
KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW cell culture replication; NS2/3; NS3/4; NS3; NS5B.
XX
OS Hepatitis C virus.
XX
PN WO200252015-A2.
XX
PD 04-JUL-2002.
XX
PF 20-DEC-2001; 2001WO-CA001843.
XX
PR 22-DEC-2000; 2000US-0257857P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Kukolj G, Pause A;
XX
DR WPI; 2002-575382/61.
DR N-PSDB; ABK88587.
XX
PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT possess enhanced transduction or replication efficiency, useful for
PT evaluating potential inhibitors of HCV replication.
XX
PS Disclosure; Page 120-129; 140pp; English.
XX
CC The invention describes a self-replicating hepatitis C virus (HCV)
CC polynucleotide molecule comprising a 5'-non translated region (NTR),
CC where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC potential inhibitors of HCV replication. The HCV RNA molecule is also
CC useful for efficiently establishing cell culture replication. The self-
CC replicating polynucleotide molecule contains a 5'-NTR, where G at
CC position 1 is substituted for A, and therefore provides an alternative to
CC existing systems comprising a self-replicating HCV RNA molecule that, in
CC conjunction with mutations in the HCV non-structural region, such as the
CC G(2042)C/R mutations, transduces and/or replicates with greater
CC efficiency. This amino acid sequence is encoded by the hepatitis C virus
CC replicon APGK12 and contains the viral protease NS2/3, protease complex
CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B
XX
SQ Sequence 2201 AA;

Query Match 95.0%; Score 1951; DB 5; Length 2201;
Best Local Similarity 93.1%; Pred. No. 7.1e-180;
Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

Qy 1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
Db 5 MAASCGGAVFVGLILLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAV 64

Qy 61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
Db 65 ILLTCAIHPELIFTITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHY 124

Qy 121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
Db 125 VQMALMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAAC 184

Qy 181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
Db 185 GDIILGLPVSARRGREIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRN 244

Qy 241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
Db 245 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 304

Qy 301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
Db 305 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 364

Qy 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
Db 365 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 397

RESULT 15
ABG30599
ID ABG30599 standard; protein; 2201 AA.
XX
AC ABG30599;
XX
DT 21-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #8.
XX

```
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 882
FT                   /label= Arg, Lys
FT   Misc-difference 1346
FT                   /note= "Wild type Leu substituted by Pro"
XX
PN   WO200252015-A2.
XX
PD   04-JUL-2002.
XX
PF   20-DEC-2001; 2001WO-CA001843.
XX
PR   22-DEC-2000; 2000US-0257857P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Kukolj G,  Pause A;
XX
DR   WPI; 2002-575382/61.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
XX
PS   Claim 3; Page; 140pp; English.
XX
CC   The invention describes a self-replicating hepatitis C virus (HCV)
CC   polynucleotide molecule comprising a 5'-non translated region (NTR),
CC   where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC   region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC   replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC   potential inhibitors of HCV replication. The HCV RNA molecule is also
CC   useful for efficiently establishing cell culture replication. The self-
CC   replicating polynucleotide molecule contains a 5'-NTR, where G at
CC   position 1 is substituted for A, and therefore provides an alternative to
CC   existing systems comprising a self-replicating HCV RNA molecule that, in
CC   conjunction with mutations in the HCV non-structural region, such as the
CC   G(2042)C/R mutations, transduces and/or replicates with greater
CC   efficiency. This amino acid sequence represents a mutant of the hepatitis
CC   C virus replicon APGK12 and contains the viral protease NS2/3, protease
CC   complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
CC   This sequence does not appear in the specification but has been created
CC   from the wild type sequence shown in ABG30580 using information given in
CC   the claims of the invention
XX
SQ   Sequence 2201 AA;

  Query Match             95.0%;  Score 1951;  DB 5;  Length 2201;
  Best Local Similarity   93.1%;  Pred. No. 7.1e-180;
  Matches  366;  Conservative   14;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| ||||||:||: |||||||||| ||| ||||||||||||||||||:
Db          5 MAASCGGAVFVGLILLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAV 64

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||:|||||| |||:|||| |||||||||||||||||| |||||||||| |||||
Db         65 ILLTCAIHPELIFTITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHY 124

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| ||||||||||||||||:|||||||||||||||||:||||| |:||||||||| .
Db        125 VQMALMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAAC 184

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||| |||||: |||||||||||||||||||||||||||||||:|
Db        185 GDIILGLPVSARRGREIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRN 244

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db        245 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 304

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 364

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||||:||||||||||
Db        365 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 397

Search completed: May  6, 2004, 09:30:45
Job time : 54.4939 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:22:36 ; Search time 12.7992 Seconds
(without alignments)
2953.573 Million cell updates/sec

Title: US-10-650-585-11
Perfect score: 2053
Sequence: 1 MAASCGGAVFIGLALLTLSP.....RGVAKAVDFIPVESMETTMR 393

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1967	95.8	3010	1	GNWVCJ	genome polyprotein
2	1950	95.0	3010	1	A45573	genome polyprotein
3	1935	94.3	3010	1	GNWVTW	genome polyprotein
4	1910	93.0	3010	1	S18030	genome polyprotein
5	1888	92.0	3010	1	GNWVTC	genome polyprotein
6	1764	85.9	3011	1	GNWVC3	genome polyprotein
7	1758	85.6	3011	1	S40770	genome polyprotein
8	1752	85.3	3011	1	GNWVCH	genome polyprotein
9	1493	72.7	3014	1	JC5620	genome polyprotein
10	1403	68.3	3033	1	JQ1303	genome polyprotein
11	1401	68.2	3033	1	GNWVJ8	genome polyprotein
12	408.5	19.9	3005	2	T08841	polyprotein - dour
13	342.5	16.7	2970	2	T08839	polyprotein - marm
14	112	5.5	692	2	H71426	hypothetical prote
15	102.5	5.0	660	1	VHWWH2	structural protein
16	101.5	4.9	3434	1	GNWVMV	genome polyprotein
17	101	4.9	564	2	S36637	signal recognition
18	101	4.9	600	2	B46642	DNA-directed DNA p
19	100.5	4.9	353	2	G87392	conserved hypothet
20	99	4.8	399	2	AH3038	conserved hypothet
21	99	4.8	399	2	C98247	hypothetical 50.8K
22	99	4.8	446	2	AF1509	conserved hypothet
23	99	4.8	451	2	H82044	C4-dicarboxylate t
24	97.5	4.7	1085	2	T03531	cobN protein homol
25	95.5	4.7	470	2	JC4098	tetracycline 6-hyd
26	94.5	4.6	904	2	A84212	hypothetical prote
27	94.5	4.6	2796	2	JC4743	fatty-acid synthas
28	94	4.6	434	2	G82728	conserved hypothet
29	93.5	4.6	418	2	H90679	probable transport
30	93.5	4.6	418	2	D85530	probable transport
31	93.5	4.6	477	2	H75026	oligopeptide abc t
32	93.5	4.6	1380	2	T18309	receptor-adenylate
33	93.5	4.6	3069	2	H70656	fatty-acid synthas
34	93	4.5	3414	1	GNWVNE	genome polyprotein
35	93	4.5	7463	2	T36248	CDA peptide synthe
36	92.5	4.5	665	2	D83252	nucleotide sugar e
37	92.5	4.5	706	2	S33761	transferrin precur
38	92.5	4.5	716	2	G83612	hypothetical prote
39	92	4.5	659	1	B44212	structural protein
40	91.5	4.5	401	1	A36961	pilin biogenesis p
41	91.5	4.5	428	2	AF0241	probable coenzyme
42	91.5	4.5	446	2	AE1150	conserved hypothet
43	91.5	4.5	3412	1	GNWVTB	genome polyprotein
44	90.5	4.4	418	2	A64763	probable transport
45	90.5	4.4	868	2	H81775	aconitate hydratas

ALIGNMENTS

RESULT 1
GNWVCJ
genome polyprotein - hepatitis C virus (strain J)
N;Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001
C;Accession: A39253; PS0086
R;Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimotc
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
A;Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients v
A;Reference number: A39253; MUID:91088550; PMID:2175903
A;Accession: A39253
A;Molecule type: genomic RNA
A;Residues: 1-3010 <KAT>
A;Cross-references: GB:D90208; NID:g221610; PIDN:BAA14233.1; PID:g221611
R;Kato, N.; Ohkoshi, S.; Shimotohno, K.
Proc. Jpn. Acad. 65B, 219-223, 1989
A;Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence varia
A;Reference number: PS0085
A;Accession: PS0086
A;Molecule type: genomic RNA
A;Residues: 2650-2707 <KA2>
A;Experimental source: Japanese isolate
C;Comment: The cleavage sites of this polyprotein have not been determined.
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2

Query Match 95.8%; Score 1967; DB 1; Length 3010;
Best Local Similarity 93.9%; Pred. No. 1.8e-153;
Matches 369; Conservative 13; Mismatches 11; Indels 0; Gaps 0;

```
Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| |||||||||| |||||||||| ||| |||||||:||||||||||||
Db        814 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFITRAEAHLQVWVPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              |||||||||||||||||||| ||||||||||:|||||||||||||||||| |||||
Db        874 ILLTCAVHPELIFDITKLLLAILGPLMVLQAGITRVPYFVRAQGLIRACMLVRKVAGGHY 933
```

```
Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              ||||||||||||||||||||||:||||||||||||||||:|||||  |:||||||||||
Db        934 VQMAFMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKLITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||||||||||||:||||||||:|  ||||||||||||||||||||||||||||||||||
Db        994 GDIISGLPVSARRGKEILLGPADSFGEQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||:|||||:|||||||||||||||||:|||||||||||||||||||||||||||||||| |
Db       1054 QVDGEVQVLSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWPA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||||||||||||||||||||:||||||||||||||:|||||||||||||||||||
Db       1114 PPGARSMTPCTCGSSDLYLVTRHADVVPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              || |||||||||||||||||||||||||||||||
Db       1174 GHVVGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206
```

RESULT 2
A45573
genome polyprotein - hepatitis C virus (strain JT)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C;Accession: A45573
R;Tanaka, T.; Kato, N.; Nakagawa, M.; Cootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,
Virus Res. 23, 39-53, 1992
A;Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: s
A;Reference number: A45573; MUID:92295714; PMID:1318627
A;Accession: A45573
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-3010 <TAN>
A;Cross-references: GB:D11168; GB:D01171; NID:g221612; PIDN:BAA01943.1; PID:g221613
A;Experimental source: HCV-JT
A;Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBIP:106207)
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

  Query Match   95.0%; Score 1950; DB 1; Length 3010;
  Best Local Similarity 93.6%; Pred. No. 4.7e-152;
  Matches 368; Conservative   9; Mismatches   16; Indels   0; Gaps   0;

```
Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||| ||:|| |||||||:||| |||||||||| ||| |||| ||:||||||||||||
Db        814 MAASCGGVVFVGLILLTLSPHYKVFLARLIWWLQYFITRAEAHLCVWVPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              |||||| |||||||||||||| |||||||| || :|||||||||||||||| |||||
Db        874 ILLTCAAHPELIFDITKLLLAILGPLMVLQAAITAMPYFVRAQGLIRACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              ||||||||||||||||||||||||||||||||||||||:|||||  ||||||||||||
Db        934 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
```

```
              |||| |||||||||||||||||: |||||||||||||:||||||||||:|||||||||
Db        994 GDIILGLPVSARRGREILLGPADSIEGQGWRLLAPITAYAQQTRGLLGCIVTSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db       1054 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWHA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              |||||||:|||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1114 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDGRGSLLSPRPVSYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||||||||||||||
Db       1174 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206
```

RESULT 3
GNWVTW
genome polyprotein - hepatitis C virus (strain Taiwan)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: A40244
R;Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992
A;Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the 5
A;Reference number: A40244; MUID:92230206; PMID:1314449
A;Accession: A40244
A;Molecule type: genomic RNA
A;Residues: 1-3010 <CHE>
A;Cross-references: GB:M84754
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

  Query Match   94.3%; Score 1935; DB 1; Length 3010;
  Best Local Similarity 91.9%; Pred. No. 8.1e-151;
  Matches 361; Conservative   17; Mismatches   15; Indels   0; Gaps   0;

```
Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| ||||||:||: |||||||||| ||| |||||||||||||||||||
Db        814 MAASCGGAVFVGLVLLTLSPHYKMFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              |||||| :|||||||||:|||| |||||||||:|::|||||||||||||||||||||||
Db        874 ILLTCAAYPELIFDITKILLAILGPLMVLQAGLTRIPYFVRAQGLIRACMLVRKAAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| |||||||||||||||||||| |||||||||||:|||||  ||||||||||||
Db        934 VQMALMKLAALTGTYVYDHLTPLQDWAHTGLRDLAVAVEPVVFSDMETKIITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||||||||: ||:||||||||||:|||||| ||||||||||||||
Db        994 GDIILGLPVSARRGREILLGPADSLEGRGWRLLAPITAYAQQTRGLFGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              |||||||||||||||||||:|||||||:|||||||||||||||||||||||||||||| |
Db       1054 QVEGEVQVVSTATQSFLATCINGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWHA 1113
```

```
Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              |  ||||:|||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db       1114 PQGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              || |||||||||||||||||:||||||||||
Db       1174 GHVVGIFRAAVCTRGVAKAVDFVPVESMETTMR 1206


RESULT 4
S18030
genome polyprotein - hepatitis C virus (isolate JK1)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Variety: isolate JK1
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001
C;Accession: S18030; S33570; A48332; S18029
R;Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A;Description: A whole genome of hepatitis C virus cDNA was isolated from a single patie
A;Reference number: S18028
A;Accession: S18030
A;Molecule type: genomic RNA
A;Residues: 1-3010 <HON>
A;Cross-references: EMBL:X61596; NID:g59478; PIDN:CAA43793.1; PID:g59479
A;Experimental source: isolate JK1 from an individual
R;Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.
Arch. Virol. 128, 163-169, 1993
A;Title: Sequence analysis of putative structural regions of hepatitis C virus isolated
A;Reference number: A48332; MUID:93119270; PMID:8380322
A;Accession: S33570
A;Molecule type: genomic RNA
A;Residues: 1-547,'T',549-621,'V',623-624,'S',626-652,'DL',655-761,'T',763-782 <HOW>
A;Cross-references: EMBL:X61591
A;Note: this sequence is inconsistent with the nucleotide translation
A;Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320
 as Trp, and TTC for residue 771 as Ser
A;Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBIP:121748)
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (As

  Query Match            93.0%;  Score 1910;  DB 1;  Length 3010;
  Best Local Similarity   91.9%;  Pred. No. 9.3e-149;
  Matches  361;  Conservative   13;  Mismatches   19;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||:|||||||||:||  |||||||||| |||||| ||||:||||||||||||
Db        814 MAASCGGAVFVGLALLTLSPHYKAFLARLIWWLQYFITRVEACLQVWVPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              |||||||| ||||||||:|||||||||||||:|:||||||||||:|||||||| |||||
Db        874 ILLTCAVHSELIFDITKILLAIFGPLMVLQAGLTRVPYFVRAQGLLRACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| |||||||||||||||:||||:|||||  ||||||:||||| ||||||||||||||
Db        934 VQMALMKLAAALTGTYVYDHLTPLRDWAHSGLRDLVVAVEPVVFSDMETKIITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||||||| |  |||||||||||||||||| |||:|||||||||
Db        994 GDIILGLPVSARRGREILLGPADGFREQGWRLLAPITAYSQQTRGLFGCIVTSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||| |||||||||||||||||||||:|||||||||||||| |||||||||||||||
Db       1054 QVEGEAQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPINQMYTNVDQDLVGWQA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              |  ||  :|||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1114 PSGAASLTPCTYGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              ||||||||||||||||||||||||||||||||
Db       1174 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206


RESULT 5
GNWVTC
genome polyprotein - hepatitis C virus
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructur
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
C;Accession: A38465
R;Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; A
J. Virol. 65, 1105-1113, 1991
A;Title: Structure and organization of the hepatitis C virus genome isolated from human c
A;Reference number: A38465; MUID:91140698; PMID:1847440
A;Accession: A38465
A;Molecule type: genomic RNA
A;Residues: 1-3010 <TAK>
A;Cross-references: EMBL:M58335; NID:g329770; PIDN:AAA72945.1; PID:g329771
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,224

  Query Match            92.0%;  Score 1888;  DB 1;  Length 3010;
  Best Local Similarity   90.8%;  Pred. No. 6.1e-147;
  Matches  357;  Conservative   13;  Mismatches   23;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||:|| |||||||||| ||||||||  || || | |||||| |||||||
Db        814 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFTTRAEADLHVWIPPLNARGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| |||||||||||||:|| |||:|||||||:|||||||||| |||||| |||||
Db        874 ILLMCAVHPELIFDITKLLIAILGPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||| |||||:|:|||||:||  ||||||||||||:||||| ||||||||||||||
Db        934 VQMAFMKLGALTGTYIYNHLTPLRDWPRAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||:|||||||: ||:| ||||||||||||||||||||||||||||||
Db        994 GDIILGLPVSARRGKEILLGPADSLEGRGLRLLAPITAYSQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||||:|||||||| |||||||||||||||||||||
Db       1054 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAAPKGPITQMYTNVDQDLVGWPK 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1114 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPF 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||||||:||||||||||
Db       1174 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 1206


RESULT 6
GNWVC3
genome polyprotein - hepatitis C virus (strain HCV-1)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001
C;Accession: A39166; PQ0403; PQ0404
R;Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
A;Title: Genetic organization and diversity of the hepatitis C virus.
A;Reference number: A39166; MUID:91172826; PMID:1848704
A;Accession: A39166
A;Molecule type: mRNA
A;Residues: 1-3011 <CHO>
A;Cross-references: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874
R;Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L
J. Gen. Virol. 73, 1131-1141, 1992
A;Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
A;Reference number: PQ0393; MUID:92268871; PMID:1316939
A;Accession: PQ0403
A;Molecule type: genomic RNA
A;Residues: 1577-1633 <CHA>
A;Cross-references: DDBJ:D10128
A;Experimental source: isolates E-b16
A;Accession: PQ0404
A;Status: preliminary
A;Molecule type: genomic RNA
A;Residues: 1577-1633 <CH2>
A;Experimental source: isolates E-b17
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

  Query Match             85.9%;  Score 1764;  DB 1;  Length 3011;
  Best Local Similarity   81.9%;  Pred. No. 1e-136;
  Matches  322;  Conservative   34;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              :|||||| | :|| ||||||||  ::  :||||| :||||| | ||||||||||||||:
Db        814 VAASCGGVVLVGLMALTLSPYYKRYISWCLWWLQYFLTRVEAQLHVWIPPLNVRGGRDAV 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| ||||| |:|||||||||:|||| :||| : ||||||| |||:| | | ||  ||||
Db        874 ILLMCAVHPTLVFDITKLLLAVFGPLWILQASLLKVPYFVRVQGLLRFCALARKMIGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||  :|| ||||||||:|||||:|||| ||||||||||||:|| || |:||||||||||
Db        934 VQMVIIKLGALTGTYVYNHLTPLRDWAHNGLRDLAVAVEPVVFSQMETKLITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              ||||:|||||||||||||||    :||||||||||:||||||||||||||||||||||||
Db        994 GDIINGLPVSARRGREILLGPADGMVSKGWRLLAPITAYAQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              |||||||:|||| |:||||||:|||||||:||||::|:| |||||: ||||||||||||
Db       1054 QVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLVGWPA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              | |:||:|||||||||||||||||||||||||||||||||||||||:||||||||||||:
Db       1114 PQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPA 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||||||::|||||
Db       1174 GHAVGIFRAAVCTRGVAKAVDFIPVENLETTMR 1206


RESULT 7
S40770
genome polyprotein - hepatitis C virus
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructur
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C;Accession: S40770; PC1285
R;Okamoto, H.
submitted to the EMBL Data Library, March 1992
A;Reference number: S40770
A;Accession: S40770
A;Molecule type: genomic RNA
A;Residues: 1-3011 <OKA>
A;Cross-references: EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g221587
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,
Jpn. J. Exp. Med. 60, 167-177, 1990
A;Title: The 5'-terminal sequence of the hepatitis C virus genome.
A;Reference number: PC1284; MUID:91013116; PMID:2170712
A;Accession: PC1285
A;Molecule type: genomic RNA
A;Residues: 1-513 <OK2>
A;Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512
A;Experimental source: isolate HC-J1
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

  Query Match             85.6%;  Score 1758;  DB 1;  Length 3011;
  Best Local Similarity   81.4%;  Pred. No. 3.2e-136;
  Matches  320;  Conservative   34;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              :|||||| | :|| ||||||||  ::  :||||| :||||| | ||:||||||||||:
Db        814 VAASCGGVVLVGLMALTLSPYYKRCISWCLWWLQYFLTRVEAQLHVWVPPLNVRGGRDAV 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| | ||| |:|||||||||: ||| :||| : ||||||| |||:| | | ||  ||||
Db        874 ILLMCVVHPTLVFDITKLLLAVLGPLWILQASLLKVPYFVRVQGLLRICALARKMVGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| :|| ||||||||:|||||:|||| ||||||||||||:|| || |:||||||||||
Db        934 VQMAIIKLGALTGTYVYNHLTPLRDWAHNGLRDLAVAVEPVVFSQMETKLITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              ||||:|||||||:||||||    :||||||||||:||||||||||||||||||||||||
Db        994 GDIINGLPVSARKGREILLGPADGMVSKGWRLLAPITAYAQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
```

```
Db       1054 QVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLVGWPA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              | ||||:||||||||||||||||||||||||||||||||||||||||:|||||||||||||:
Db       1114 PQGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPA 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              || |||||||||||||||||||||||:|||||
Db       1174 GHVVGIFRAAVCTRGVAKAVDFIPVESLETTMR 1206


RESULT 8
GNWVCH
genome polyprotein - hepatitis C virus (strain H)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: A36814; A41546
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A;Description: Genomic structure of the human prototype strain H of hepatitis C virus: c
A;Reference number: A36814
A;Accession: A36814
A;Molecule type: genomic RNA
A;Residues: 1-3011 <INC>
A;Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A;Title: Genomic structure of the human prototype strain H of hepatitis C virus: compari
A;Reference number: A41546; MUID:92052256; PMID:1658800
A;Contents: annotation
A;Note: neither amino acid nor nucleotide sequence is given
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

  Query Match             85.3%;  Score 1752;  DB 1;  Length 3011;
  Best Local Similarity   81.4%;  Pred. No. 1e-135;
  Matches  320;  Conservative   36;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              :|||||| | :||  |||||||  ::  :|||| :|||| | ||:|||||||||||:
Db        814 VAASCGGVVLVGLMALTLSPYYKRYISWCMWWLQYFLTRVEAQLHVWVPPLNVRGGRDAV 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||| ||| |:|||||||||||||| :||| : ||||||| |||:| | | || |||||
Db        874 ILLTCVVHPALVFDITKLLLAIFGPLWILQASLLKVPYFVRVQGLLRICALARKIAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| :|| ||||| ||:|| ||:|||| ||||||||||||:|| || |:||||||||||
Db        934 VQMAIIKLGALTGTCVYNHLAPLRDWAHNGLRDLAVAVEPVVFSRMETKLITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              ||||:|||||||||:||||||||    :||||||||||:|||||||||||||||||||||
Db        994 GDIINGLPVSARRGQEILLGPADGMVSKGWRLLAPITAYAQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              |||||||:|||||:|||||:|||||||:||||::|:| ||||: | ||||||||||| |
```

```
Db       1054 QVEGEVQIVSTATQTFLATCINGVCWTVYHGAGTRTIASPKGPVIQTYTNVDQDLVGWPA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              | |:||:|||||||||||||||||||||||||||||||||||||:|||||||||||||:
Db       1114 PQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPT 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||:|||||||||||||||||::|||||
Db       1174 GHAVGLFRAAVCTRGVAKAVDFIPVENLETTMR 1206


RESULT 9
JC5620
genome polyprotein - hepatitis C virus (isolate EUH1480)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructur
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C;Accession: JC5620
R;Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
Biochem. Biophys. Res. Commun. 236, 44-49, 1997
A;Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant g
A;Reference number: JC5620; MUID:97366593; PMID:9223423
A;Accession: JC5620
A;Molecule type: mRNA
A;Residues: 1-3014 <CHA>
A;Cross-references: GB:Y13184
A;Experimental source: genotype 5a, which predominates in South Africa
A;Note: the translation of the nucleotide sequence is not complete in this paper
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;384-408/Region: hypervariable #status predicted
F;390-730/Product: nonstructural protein NS1 #status predicted <NS1>
F;731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
F;1008-1616/Product: hepacivirin #status predicted <NS3>
F;1231-1238/Region: nucleotide-binding motif A (P-loop)
F;1313-1318/Region: nucleotide-binding motif B
F;1317-1320/Region: DEXH motif
F;1617-1863/Product: nonstructural protein NS4a #status predicted <N4A>
F;1864-2014/Product: nonstructural protein NS4b #status predicted <N4B>
F;2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
F;2210-2249/Region: interferon sensitivity determining #status predicted

  Query Match             72.7%;  Score 1493;  DB 1;  Length 3014;
  Best Local Similarity   68.8%;  Pred. No. 2.3e-114;
  Matches  267;  Conservative   56;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          6 GGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTC 65
              |  | | | : ||:| ||  :  ::|||| | |||| : ||:||| |:|||||:|:|||
Db        820 GCLVLIVLTIFTLTPGYKKKVVLVMWWLQYFIARVEAIIHVWVPPLQVKGGRDAVIMLTC 879

Qy         66 AVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAF 125
                || | |:|||:|  | ||| :||  :||||||:|| |:|  |:|  :    | ||| |
Db        880 LFHPALGFEITKILFGILGPLYLLQHSLTKVPYFLRARALLRLCLLAKHLVYGKYVQAAL 939

Qy        126 MKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIIS 185
              : |  |||||:|||| |::||| :|||:| || ||::|| || |:||||||||||:|::
Db        940 LHLGRLTGTYIYDHLAPMKDWAASGLRELTVATEPIVFSAMETKVITWGADTAACGNILA 999

Qy        186 GLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGE 245
              |||||||||| |||||: :  ||||||||||:|||||:|| |: |||||||: |||
Db       1000 VLPVSARRGREIFLGPADDIKTSGWRLLAPITAYAQQTRGVLGAIVLSLTGRDKNEAEGE 1059

Qy        246 VQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGAR 305
              || :|||||:||  |:||| ||:|||||||||||||: |||||||:||||| :|||
Db       1060 VQFLSTATQTFLGICINGVMWTLFHGAGSKTLAGPKGPVVQMYTNVDKDLVGWPSPPGKG 1119

Qy        306 SMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVG 365
              |:| |||||:||||||||||||| |||||:| |||||||:||||||||::||||| ||
Db       1120 SLTRCTCGSADLYLVTRHADVIPARRRGDTRASLLSPRPISYLKGSSGGPIMCPSGHVVG 1179
```

```
Qy        366 IFRAAVCTRGVAKAVDFIPVESMETTMR 393
              :|||||||||||||::|:|||::|||||
Db       1180 VFRAAVCTRGVAKALEFVPVENLETTMR 1207
```

RESULT 10
JQ1303
genome polyprotein - hepatitis C virus (isolate HC-J6)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
C;Accession: JQ1303
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.
J. Gen. Virol. 72, 2697-2704, 1991
A;Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a hum
A;Reference number: JQ1303; MUID:92044440; PMID:1658196
A;Accession: JQ1303
A;Molecule type: genomic RNA
A;Residues: 1-3033 <OKA>
A;Cross-references: GB:D00944; NID:g221650; PIDN:BAA00792.1; PID:g221651
A;Experimental source: isolate HC-J6 from a Japanese individual
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transm
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F;1011-1619/Product: hepacivirin #status predicted <NS3>
F;1316-1321/Region: nucleotide-binding motif B
F;1320-1323/Region: DEXH motif
F;1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>
F;1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>
F;2018-3033/Product: nonstructural protein NS5 #status predicted <NO5>
F;196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28

Query Match 68.3%; Score 1403; DB 1; Length 3033;
Best Local Similarity 64.9%; Pred. No. 6.2e-107;
Matches 252; Conservative 58; Mismatches 78; Indels 0; Gaps 0;

```
Qy          6 GGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTC 65
              | |: : : | ||:| || ||:| :||| ||:|  || :| | ||: ||||||  ||
Db        823 GAALLVLITLFTLTPGYKTLLSRFLWWLCYLLTLAEAMVQEWAPPMQVRGGRDGIIWAVA 882

Qy         66 AVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAF 125
                 | ::||||| |||: ||  :|:  :|:|||||||  |:| | :||  ||| ||||
Db        883 IFCPGVVFDITKWLLAVLGPAYLLKGALTRVPYFVRAHALLRMCTMVRHLAGGRYVQMVL 942

Qy        126 MKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIIS 185
              : |   ||||:||||||: |||  ||||||||||:||| || |:| |||:|||||||:
Db        943 LALGRWTGTYIYDHLTPMSDWAAAGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILH 1002

Qy        186 GLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGE 245
              ||||||| |||:|||||| :  :|| ||||||||:|||||||| |: |:|||| :  ||
Db       1003 GLPVSARLGREVLLGPADGYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGE 1062

Qy        246 VQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGAR 305
              :||:|| ||||| | ::|| |||:||||:||||| :||:||||:: : ||||| :||| :
Db       1063 IQVLSTVTQSFLGTTISGVLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTK 1122

Qy        306 SMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVG 365
              |: ||||||: ||||||:||||| ||||| ||:||||||:| |||||||:||| |||||
Db       1123 SLEPCTCGAVDLYLVTRNADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVG 1182

Qy        366 IFRAAVCTRGVAKAVDFIPVESMETTMR 393
              :||||||:|||||::||||||:::  |
Db       1183 VFRAAVCSRGVAKSIDFIPVETLDIVTR 1210
```

RESULT 11
GNWVJ8

genome polyprotein - hepatitis C virus (strain HC-J8)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: A40250; PQ0397; PQ0559
R;Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.;
Virology 188, 331-341, 1992
A;Title: Full-length sequence of a hepatitis C virus genome having poor homology to repo
A;Reference number: A40250; MUID:92230232; PMID:1314459
A;Accession: A40250
A;Molecule type: genomic RNA
A;Residues: 1-3033 <OKA>
A;Cross-references: GB:D10988; GB:D01221; NID:g221608; PIDN:BAA01761.1; PID:g221609
R;Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L.
J. Gen. Virol. 73, 1131-1141, 1992
A;Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
A;Reference number: PQ0393; MUID:92268871; PMID:1316939
A;Accession: PQ0397
A;Molecule type: genomic RNA
A;Residues: 2678-2754 <CHA>
A;Cross-references: DDBJ:D10134
A;Experimental source: isolate E-b12
R;Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno,
Biochem. Biophys. Res. Commun. 181, 279-285, 1991
A;Title: Distribution of plural HCV types in Japan.
A;Reference number: PQ0554; MUID:92068204; PMID:1720309
A;Accession: PQ0559
A;Molecule type: mRNA
A;Residues: 2678-2729 <KAT>
A;Cross-references: GB:D10562; GB:D90518; NID:g221523; PIDN:BAA01418.1; PID:g221524
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F;1011-1619/Product: hepacivirin #status predicted <NS3>
F;1234-1241/Region: nucleotide-binding motif A (P-loop)
F;1316-1321/Region: nucleotide-binding motif B
F;1320-1323/Region: DEXH motif
F;1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>
F;1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>
F;2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,235

Query Match 68.2%; Score 1401; DB 1; Length 3033;
Best Local Similarity 63.4%; Pred. No. 9e-107;
Matches 246; Conservative 67; Mismatches 75; Indels 0; Gaps 0;

```
Qy          6 GGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTC 65
              | |: : ::: ||:| ||:||:| :||| |::  || :| |:||| |||||| || :
Db        823 GLAILVIISIFTLTPAYKILLSRSVWWLSYMLVLAEAQIQQWVPPLEVRGGRDGIIWVAV 882

Qy         66 AVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAF 125
                :|| |:|::|| |||| ||  :|:| : ::||||||  |:| | ||:  ||  |:||
Db        883 ILHPRLVFEVTKWLLAILGPAYLLKASLLRIPYFVRAHALLRVCTLVKHLAGARYIQMLL 942

Qy        126 MKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIIS 185
              : :   ||||:||||:||  ||  ||||||:|||||:|| || |:| |||:| |||||:
Db        943 ITIGRWTGTYIYDHLSPLSTWAAQGLRDLAIAVEPVVFSPMEKKVIVWGAETVACGDILH 1002

Qy        186 GLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGE 245
              ||||||| |||:|||||| :  :||:|||||||:|||||||| |: |||||||:  |:
Db       1003 GLPVSARLGREVLLGPADGYTSKGWKLLAPITAYTQQTRGLLGAIVVSLTGRDKNEQAGQ 1062

Qy        246 VQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGAR 305
              |||:|: ||:|| | ::|| |||:||||:|||||||||:|||||: : ||||| :|:|| :
Db       1063 VQVLSSVTQTFLGTSISGVLWTVYHGAGNKTLAGPKGPVTQMYTSAEGDLVGWPSPPGTK 1122

Qy        306 SMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVG 365
              |: ||||||: |||||||:||||||||: | ||:|||||||:| ||||||||:||  |||||
Db       1123 SLEPCTCGSVDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVG 1182
```

```
Db       1123 SLDPCTCGAVDLYLVTRNADVIPVRRKDDRRGALLSPRPLSTLKGSSGGPVLCSRGHAVG 1182

Qy        366 IFRAAVCTRGVAKAVDFIPVESMETTMR 393
              :|||||| ||||:::|||||||::   |
Db       1183 LFRAAVCARGVAKSIDFIPVESLDVATR 1210


RESULT 12
T08841
polyprotein - douroucouli hepatitis GB virus A
C;Species: douroucouli hepatitis GB virus A
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C;Accession: T08841
R;Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
J. Gen. Virol. 79, 41-45, 1998
A;Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
A;Reference number: Z16486; MUID:98120818; PMID:9460920
A;Accession: T08841
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-3005 <ERK>
A;Cross-references: EMBL:AF023425; NID:g2828599; PIDN:AAC40502.1; PID:g2828600
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: polyprotein

  Query Match             19.9%;  Score 408.5;  DB 2;  Length 3005;
  Best Local Similarity   31.3%;  Pred. No. 6.5e-25;
  Matches  111;  Conservative   57;  Mismatches  154;  Indels   33;  Gaps    9;

Qy         54 RGGRD----AIILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRAC 109
              |||||    |: :    : | :   :  | |       | | :  : | :    :: |
Db        813 RGGRDWRVTAVWVAAGIIFPREVVRCSAALTA-------LAALLDSIDYLLETL-ILTAA 864

Qy        110 MLVRKA--------AGGHYVQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPV 161
              | |           |    :  ||::     |  :: |   :   | | | || ||:|||
Db        865 QPARAARLLDSLTFLGDADLTRAFVRRLERRGVTLFQHCGQVSXGAAAILXDLGVALEPV 924

Qy        162 IFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLG--PADNFEGQGWRLLAPITAY 219
              :  :  |:   | | ||| : |||| |||| |:|:|  |:      |:   ||:
Db        925 SVTARDCYIVRDAARTLACGQRVEGLPVVARRGDEVLVGVFPSVRALPPGFVPTAPVVV- 983

Qy        220 SQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAG 279
              |:  |   :  ||: |||: : || : |: |:|   : |||||| :| |||: ::||||
Db        984 MQRGLGFFSVVKTSMLGRDEREHEGSIVVLGTSTTRSMGTCVNGVMYTTFHGSNARTLAG 1043

Qy        280 PKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSL 339
              | ||:   : :   |: :  | ||  : || ||:  :: :    |       |       |
Db       1044 PVGPVNCRWWSPSDDVAVYPLPSGASCLEPCKCGTQSVWCIRN--DGALCHGRLSKLVEL 1101

Qy        340 LSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGV-------AKAVDFIPVES 387
              |  :|  :|||| |:||  || ||:   :|  |||         |  : :| :|
Db       1102 DLPTEISDFRGSSGSPILCDEGHVVGMM-VSVLHRGVKVTGVRYVKPWETLPKDS 1155


RESULT 13
T08839
polyprotein - marmoset hepatitis GB virus A
C;Species: marmoset hepatitis GB virus A
C;Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 17-Nov-2000
C;Accession: T08839
R;Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
J. Gen. Virol. 79, 41-45, 1998
A;Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
A;Reference number: Z16486; MUID:98120818; PMID:9460920
A;Accession: T08839
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: genomic RNA
A;Residues: 1-2970 <ERK>
A;Cross-references: EMBL:AF023424; NID:g2828597; PIDN:AAC40501.1; PID:g2828598
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: polyprotein

  Query Match             16.7%;  Score 342.5;  DB 2;  Length 2970;
  Best Local Similarity   28.9%;  Pred. No. 1.8e-19;
  Matches  103;  Conservative   56;  Mismatches  133;  Indels   65;  Gaps   12;

Qy         78 LLLAIFGP--------LMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLA 129
              |: | | |          :|  |   |  ::    |: :  ||| |     :  |  |||
Db        816 LVAAWFWPREIAAVCVAFILGFGFFDVVDYILEVVLVSSPNLVRLARVLDSLVAAGDKLA 875

Qy        130 ALTGTYVYDHLTPLQD--WAHAG---------LRDLAVAVEPVIFSDMEVKIITWGADTA 178
                 |:: : |        :||||          ||     |:|||      :  ::    | |
Db        876 T---TWLVEKLRKRNCFLYAHAGQVTRRTAEQLRQWGFALEPVAVHPEDCAMVRDAARTL 932

Qy        179 ACGDIISGLPVSARRGREILLGPADNFEGQGWRL------LAPITAYSQQTRGLLGCIIT 232
              :||  : | || |||| |:|:|  :          | |          ||:  :    :|  | : |
Db        933 SCGQSVHGKPVVARRGDEVLIGVLNGV----WELPPGFVPTAPVVVH-HHGKGFFGVVKT 987

Qy        233 SLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVD 292
              |:|| |: :  | | |: |:|    : ||||||  :| :||: ::|||    ||:    : :
Db        988 SMTGWDETEHVGNVVVLGTSTTRSMGTCVNGVMYTTYHGSNARTLAAQMGPVNSRWWSAS 1047

Qy        293 QDLVGWQAPPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLS---------- 341
               |:  :  | ||: : || |     ::::            | |  |:|
Db       1048 DDVAVYPLPVGAKCLEPCKCQPQGVWVI----------RND--GALCHGTLGRTVELDL 1094

Qy        342 PRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRG-------VAKAVDFIPVESMETT 391
              |  :  :|||| |:||  |||||:   :|  ||           |  : :| |::  |
Db       1095 PAELCDFRGSSGSPILCDEGHAVGML-ISVLHRGSRVTGIRYTKPWETLPREAITHT 1150


RESULT 14
H71426
hypothetical protein - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
A;Variety: columbia
C;Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998
C;Accession: H71426
R;Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirks
 P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.; Giele
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A;Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech,
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ansc
C.; Chalwatzis, N.
A;Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal:
A;Reference number: A71400; MUID:98121113; PMID:9461215
A;Accession: H71426
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-692 <BEV>
A;Cross-references: GB:Z97340; NID:g2244950; PID:e327492; PID:g2244965
C;Genetics:
A;Map position: 4COP9-4G3845

  Query Match              5.5%;  Score 112;  DB 2;  Length 692;
  Best Local Similarity   24.1%;  Pred. No. 0.3;
  Matches   97;  Conservative   45;  Mismatches  135;  Indels  126;  Gaps   23;

Qy         26 LARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIFDITKLLLAIFGP 85
              :| | : : ||: |      |   | :|  | : | |       ||:| |:|:||   |
Db         83 IADLAFGINYLMRR-----QGNFPTASVYAGSNCIEL----KGPEIIMDLTELLR--FLT 131

Qy         86 LMVLQAGITKVPY--FVRAQGLIRACMLVR--KAAGGHYVQMAFM-------------- 126
              | :|     :| |:  |: : |     :|::  ||  || :| ||
Db        132 LCML---FSKKPFAVFLESAGYTHEDVLLQKPKAGVGHIMQPAFTIIRDTNSKCILLLIR 188

Qy        127 -------KLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAA 179
                    | | ||  |  | : | |     || :| :                      |
Db        189 GTHSIKDTLTAATGAVVPFHHSVLHD---GGLSNLVLGY------------------AH 226

Qy        180 CGDIISGLPVSARRGREILLGPADNFEGQGW--RLLAPITAYSQQTRGLLGCIITSLTGR 237
              ||       |:| |                  |  :|  |             |:: :|
Db        227 CG------MVAAAR---------------WIAKLSVP-------------CLLKAL--- 248

Qy        238 DKNQVEGEVQVVSTA----TQSFLATCV-------NGVCWTVFHGAGSKTLAGPKGP--I 284
```

```
Db        249 DENP-SFKVQIVGHSLGGGTASLLTYILREQKEFASATCFTFAPGTPNLMINGESGKHFI 307

Qy        285 TQMYTNVDQDLV---GWQAPPGARSMTPCTCGSSDLYLVTRHADVIP-VRRRGDSRGSLL 340
Db        308 TTIING--SDLVPTFSASSVDDLRSEVTSSSWSNDLRDQVEHTRVLSVVYRSATAIGSRL 365

Qy        341 SPRPVSYLKGSSGGPLLCP--SGHAVGIFRAAVCTRGVAKAVD 381
Db        366 PSIASAKAKVAGAGAILRPVSSGTQVAAFLVNGC--GKIKCID 406


RESULT 15
VHWWH2
structural protein 2 precursor - hepatitis E virus (strain Burma)
C;Species: hepatitis E virus
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 23-Jul-1999
C;Accession: C40778
R;Tam, A.W.; Smith, M.M.; Guerra, M.E.; Huang, C.C.; Bradley, D.W.; Fry, K.E.; Reyes, G.
Virology 185, 120-131, 1991
A;Title: Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length vi
A;Reference number: A40778; MUID:92024067; PMID:1926770
A;Accession: C40778
A;Molecule type: genomic RNA
A;Residues: 1-660 <TAM>
A;Cross-references: GB:M73218; NID:g330023; PIDN:AAA45736.1; PID:g330026
A;Note: the authors translated the codon CGC for residue 2 as Ala
C;Superfamily: hepatitis E virus structural protein 2
C;Keywords: structural protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-660/Product: structural protein 2 #status predicted <SP2>

  Query Match             5.0%;  Score 102.5;  DB 1;  Length 660;
  Best Local Similarity   19.8%;  Pred. No. 1.7;
  Matches   82;  Conservative   52;  Mismatches  130;  Indels  151;  Gaps   19;

Qy         76 TKLLL--AIFGPLMVLQAG------ITKVPYFVR---AQGLIRACMLVRKAAGGHYVQMA 124
Db        151 TNLVLYAAPLSPLLPLQDGTNTHIMATEASNYAQYRVARATIRYRPLVPNAVGGYAISIS 210

Qy        125 FMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDII 184
Db        211 FWPQTTTTPTSV--------DMNSITSTDVRILVQPGIASELVI---------------- 246

Qy        185 SGLPVSARRGREILLGPAD--NFEGQGWRLLAPI-TAYSQQTRGLL-----GCIITSLTG 236
Db        247 ---------------PSERLHYRNQGWRSVETSGVAEEEATSGLVMLCIHGSLVNSYTN 290

Qy        237 ----------------------------------------RDKNQVEGEVQVVSTATQSFL 257
Db        291 TPYTGALGLLDFALELEFRNLTPGNTNTRVSRYSSTARHRLRRGADGTAELTTTAATRFM 350

Qy        258 A----TCVNGV-------CWTVFH---------------GAG-----SKTLAGPKG-PIT 285
Db        351 KDLYFTSTNGVGEIGRGIALTLFNLADTLLGGLPTELISSAGGQLFYSRPVVSANGEPTV 410

Qy        286 QMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLV---TRHADVIPVRRRGDSRG-SLLS 341
Db        411 KLYTSVENA----QQDKGIAIPHDIDLGESRVVIQDYDNQHEQDRPTPSPAPSRPFSVLR 466

Qy        342 PRPVSYLK-----------GSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPV 385
Db        467 ANDVLWLSLTAAEYDQSTYGSSTGPVYV--SDSVTLVNVATGAQAVARSLDWTKV 519


Search completed: May  6, 2004, 09:37:17
Job time : 14.7992 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:09:55 ; Search time 8.20459 Seconds
(without alignments)
2494.160 Million cell updates/sec

Title: US-10-650-585-11
Perfect score: 2053
Sequence: 1 MAASCGGAVFIGLALLTLSP.....RGVAKAVDFIPVESMETTMR 393

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1967	95.8	3010	1	POLG_HCVJA	P26662 h genome po
2	1950	95.0	3010	1	POLG_HCVJT	Q00269 h genome po
3	1935	94.3	3010	1	POLG_HCVTW	P29846 h genome po
4	1888	92.0	3010	1	POLG_HCVBK	P26663 h genome po
5	1764	85.9	3011	1	POLG_HCV1	P26664 h genome po
6	1752	85.3	3011	1	POLG_HCVH	P27958 h genome po
7	1403	68.3	3033	1	POLG_HCVJ6	P26660 h genome po
8	1401	68.2	3033	1	POLG_HCVJ8	P26661 h genome po
9	102.5	5.0	660	1	VST2_HEVBU	P29326 hepatitis e
10	102.5	5.0	660	1	VST2_HEVPA	P33426 hepatitis e
11	101.5	4.9	1780	1	POLG_MVEV	P05769 m genome po
12	101	4.9	564	1	SR5C_ARATH	P37107 arabidopsis
13	101	4.9	600	1	DPO2_MOUSE	P33611 mus musculu
14	94	4.6	444	1	YA47_XYLFA	Q9pei1 xylella fas
15	93.5	4.6	1380	1	CYAA_LEIDO	Q27675 leishmania
16	93	4.5	434	1	TOLB_CHLTE	Q8keq0 chlorobium
17	92.5	4.5	706	1	TRFE_HORSE	P27425 equus cabal
18	92	4.5	444	1	Y327_XYLFT	Q87ei0 xylella fas
19	92	4.5	659	1	VST2_HEVME	Q03500 hepatitis e
20	92	4.5	3414	1	POLG_TBEVW	P14336 t genome po
21	91.5	4.5	401	1	PILC_PSEPU	P36641 pseudomonas
22	91.5	4.5	485	1	VST2_HEVRH	Q00270 hepatitis e
23	91.5	4.5	660	1	VST2_HEVMY	Q04611 hepatitis e
24	91.5	4.5	3412	1	POLG_TBEVS	P07720 t genome po
25	90.5	4.4	403	1	MHPT_ECOLI	P77589 escherichia
26	90.5	4.4	3414	1	POLG_TBEVH	Q01299 t genome po
27	89.5	4.4	961	1	ATCU_YERPE	Q8zca7 yersinia pe
28	88.5	4.3	355	1	MURG_NEIMA	Q9jsz7 neisseria m
29	88.5	4.3	3432	1	POLG_JAEVJ	P32886 j genome po
30	88	4.3	441	1	MDHP_MESCR	Q05145 mesembryant
31	87.5	4.3	3432	1	POLG_JAEV1	P27395 j genome po
32	87.5	4.3	3432	1	POLG_JAEV5	P19110 j genome po
33	86.5	4.2	347	1	MDHM_EUCGU	P46487 eucalyptus
34	86.5	4.2	470	1	NRAM_IAWHM	P05803 influenza a
35	86	4.2	338	1	GALE_NEIGO	Q05026 neisseria g
36	86	4.2	433	1	DCUA_WOLSU	O34245 wolinella s
37	86	4.2	790	1	RIR1_HSVEB	P28846 equine herp
38	85.5	4.2	355	1	YP71_MYCTU	Q50650 mycobacteri
39	85	4.1	470	1	NRAM_IATRA	P03472 influenza a
40	85	4.1	730	1	HELS_METMA	Q8pzr7 methanosarc
41	85	4.1	854	1	PWP2_SCHPO	Q9c1x1 schizosacch
42	85	4.1	3313	1	CLR3_RAT	O88278 rattus norv
43	84.5	4.1	453	1	NHAC_BACSU	O07553 bacillus su
44	84.5	4.1	1705	1	PTPV_MOUSE	P70289 mus musculu
45	84	4.1	305	1	OPPC_BACSU	P24139 bacillus su

ALIGNMENTS

```
RESULT 1
POLG_HCVJA
ID   POLG_HCVJA     STANDARD;      PRT;  3010 AA.
AC   P26662;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate Japanese) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91088550; PubMed=2175903;
RA   Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA   Sugimura T., Shimotohno K.;
RT   "Molecular cloning of the human hepatitis C virus genome from
RT   Japanese patients with non-A, non-B hepatitis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RN   [2]
RP   DISCUSSION OF SEQUENCE.
RX   MEDLINE=91192160; PubMed=1849488;
RA   Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraiso K.,
RA   Ohkoshi S., Shimotohno K.;
RT   "Molecular structure of the Japanese hepatitis C viral genome.";
RL   FEBS Lett. 280:325-328(1991).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D90208; BAA14233.1; -.
```

```
DR   PIR; A39253; GNWVCJ.
DR   HSSP; P26663; 1JXP.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT   CHAIN       730   1006       NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3010       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
FT   SITE       1316   1319       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2041   2041       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2077   2077       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2240   2240       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2788   2788       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3010 AA;  327017 MW;  AA993794F46DB185 CRC64;

 Query Match             95.8%;  Score 1967;  DB 1;  Length 3010;
  Best Local Similarity   93.9%;  Pred. No. 1.8e-151;
  Matches  369;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| |||||||||| |||||||||| ||| |||||||:||||||||||||
Db        814 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFITRAEAHLQVWVPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              |||||||||||||||||||| ||||||||||:|||||||||||||||||| |||||
Db        874 ILLTCAVHPELIFDITKLLLAILGPLMVLQAGITRVPYFVRAQGLIRACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              ||||||||||||||||||||:||||||||||||||:||||| |:||||||||||
Db        934 VQMAFMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKLITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||||||||||||:||||||||:|  |||||||||||||||||||||||||||||||||
Db        994 GDIISGLPVSARRGKEILLGPADSFGEQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||:|||||:||||||||||||||||||:||||||||||||||||||||||||||||||| |
Db       1054 QVDGEVQVLSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWPA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||||||||||||||||||||:||||||||||||||:|||||||||||||||||||
Db       1114 PPGARSMTPCTCGSSDLYLVTRHADVVPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              || |||||||||||||||||||||||||||||||
Db       1174 GHVVGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206


RESULT 2
POLG_HCVJT
ID   POLG_HCVJT     STANDARD;      PRT;  3010 AA.
AC   Q00269;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate HC-JT) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=31642;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92295714; PubMed=1318627;
RA   Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA   Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT   "Molecular cloning of hepatitis C virus genome from a single Japanese
RT   carrier: sequence variation within the same individual and among
RT   infected individuals.";
RL   Virus Res. 23:39-53(1992).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D11168; BAA01943.1; -.
DR   PIR; A45573; A45573.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT   CHAIN       730   1006       NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3010       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
FT   SITE       1316   1319       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2041   2041       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2077   2077       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2240   2240       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2529   2529       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2788   2788       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3010 AA;  326573 MW;  94A1C77435D642BB CRC64;

  Query Match             95.0%;  Score 1950;  DB 1;  Length 3010;
  Best Local Similarity   93.6%;  Pred. No. 4.2e-150;
  Matches  368;  Conservative    9;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||| ||:|| |||||:||| |||||||||| ||| |||| ||:|||||||||||||
Db        814 MAASCGGVVFVGLILLTLSPHYKVFLARLIWWLQYFITRAEAHLCVWVPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              |||||| |||||||||||||||| |||||||| || :|||||||||||||||||| |||||
Db        874 ILLTCAAHPELIFDITKLLLAILGPLMVLQAAITAMPYFVRAQGLIRACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              ||||||||||||||||||||||||||||||||||||||||:||||| ||||||||||||
Db        934 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||||||: ||||||||||||||:||||||||||:||||||||||
Db        994 GDIILGLPVSARRGREILLGPADSIEGQGWRLLAPITAYAQQTRGLLGCIVTSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db       1054 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWHA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              |||||:|||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       1114 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDGRGSLLSPRPVSYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||||||||||||||
Db       1174 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206


RESULT 3
POLG_HCVTW
ID   POLG_HCVTW     STANDARD;      PRT;  3010 AA.
AC   P29846;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate Taiwan) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=31645;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92230206; PubMed=1314449;
RA   Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
RT   "The Taiwanese hepatitis C virus genome: sequence determination and
RT   mapping the 5' termini of viral genomic and antigenomic RNA.";
RL   Virology 188:102-113(1992).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
```

```
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       (RNA)(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M84754; -; NOT_ANNOTATED_CDS.
DR   PIR; A40244; GNWVTW.
DR   PDB; 1N64; 25-FEB-03.
DR   PDB; 1NS3; 08-APR-98.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CORE PROTEIN (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT   CHAIN       730   1006       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3010       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE    1083   1083      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1107   1107      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1165   1165      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND     1230   1237      ATP (POTENTIAL).
FT   SITE        1316   1319      DECH BOX.
FT   CARBOHYD     196    196      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     209    209      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     233    233      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     234    234      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     250    250      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     305    305      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     417    417      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     423    423      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     430    430      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     448    448      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     532    532      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     540    540      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     556    556      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     576    576      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     623    623      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     645    645      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2041   2041      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2077   2077      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2240   2240      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2529   2529      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2788   2788      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3010 AA;  327047 MW;  AAD267D55CDFE215 CRC64;
```

Query Match 94.3%; Score 1935; DB 1; Length 3010;
Best Local Similarity 91.9%; Pred. No. 7e-149;
Matches 361; Conservative 17; Mismatches 15; Indels 0; Gaps 0;

```
Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| |||||||:||: |||||||||| ||| ||||||||||||||||||||
Db        814 MAASCGGAVFVGLVLLTLSPHYKMFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              |||||| :|||||||||:|||| |||||||||:|::|||||||||||||||||||||||||
Db        874 ILLTCAAYPELIFDITKILLAILGPLMVLQAGLTRIPYFVRAQGLIRACMLVRKAAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| |||||||||||||||||||||| |||||||||||:||||| ||||||||||||
Db        934 VQMALMKLAALTGTYVYDHLTPLQDWAHTGLRDLAVAVEPVVFSDMETKIITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||||||: ||:||||||||||:|||||| ||||||||||||||||
Db        994 GDIILGLPVSARRGREILLGPADSLEGRGWRLLAPITAYAQQTRGLFGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||:|||||||:|||||||||||||||||||||||||||||| |
Db       1054 QVEGEVQVVSTATQSFLATCINGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWHA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              | ||||:||||||||||||||||||||||||||||||||||:||||||||||||||||
Db       1114 PQGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              || |||||||||||||||||:||||||||||
Db       1174 GHVVGIFRAAVCTRGVAKAVDFVPVESMETTMR 1206
```

```
RESULT 4
POLG_HCVBK
ID   POLG_HCVBK     STANDARD;      PRT;  3010 AA.
AC   P26663;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate BK) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11105;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91140698; PubMed=1847440;
RA   Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RA   Onishi E., Andoh T., Yoshida I., Okayama H.;
RT   "Structure and organization of the hepatitis C virus genome isolated
RT   from human carriers.";
RL   J. Virol. 65:1105-1113(1991).
RN   [2]
RP   SEQUENCE OF 1487-1500.
RX   MEDLINE=96235224; PubMed=8647104;
RA   Borowski P., Heiland M., Oehlmann K., Becker B., Korneteky L.;
RT   "Non-structural protein 3 of hepatitis C virus inhibits
RT   phosphorylation mediated by cAMP-dependent protein kinase.";
RL   Eur. J. Biochem. 237:611-618(1996).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX   MEDLINE=97015088; PubMed=8861916;
RA   Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
RA   Moomaw E.W., Adachi T., Hostomska Z.;
RT   "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RT   trypsin-like fold and a structural zinc binding site.";
RL   Cell 87:331-342(1996).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX   MEDLINE=98227846; PubMed=9568891;
RA   Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
RA   Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
RT   "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
RT   virus: a 2.2-A resolution structure in a hexagonal crystal form.";
RL   Protein Sci. 7:837-847(1998).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M58335; AAA72945.1; -.
DR   PIR; A38465; GNWVTC.
DR   PDB; 1A1Q; 25-MAR-98.
DR   PDB; 1JXP; 14-JAN-98.
DR   PDB; 1NS3; 08-APR-98.
DR   PDB; 1C2P; 15-NOV-00.
DR   PDB; 1CSJ; 08-NOV-99.
DR   PDB; 1GX5; 09-APR-02.
DR   PDB; 1GX6; 10-APR-02.
DR   PDB; 1QUV; 26-JUN-00.
DR   PDB; 8OHM; 20-APR-99.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT   CHAIN       730   1006       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3010       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM.
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM.
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM.
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
FT   SITE       1316   1319       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2041   2041       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2077   2077       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2240   2240       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2529   2529       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2788   2788       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   STRAND     1031   1035
FT   HELIX      1039   1047
FT   STRAND     1050   1050
FT   STRAND     1059   1063
FT   STRAND     1068   1074
FT   TURN       1075   1076
FT   STRAND     1077   1081
FT   HELIX      1082   1085
FT   TURN       1086   1087
FT   STRAND     1090   1092
FT   TURN       1093   1094
FT   STRAND     1095   1097
FT   STRAND     1101   1103
FT   TURN       1104   1107
```

```
FT   STRAND      1108   1112
FT   STRAND      1120   1120
FT   STRAND      1122   1122
FT   STRAND      1129   1133
FT   TURN        1135   1136
FT   STRAND      1139   1144
FT   STRAND      1149   1157
FT   HELIX       1158   1161
FT   TURN        1162   1163
FT   TURN        1165   1166
FT   STRAND      1168   1171
FT   TURN        1172   1174
FT   STRAND      1175   1186
FT   TURN        1187   1188
FT   STRAND      1189   1197
FT   HELIX       1198   1202
FT   TURN        1203   1204
FT   STRAND      1680   1688
SQ   SEQUENCE   3010 AA;  327189 MW;  F8422D5ECCFDFD9C CRC64;

  Query Match             92.0%;  Score 1888;  DB 1;  Length 3010;
  Best Local Similarity   90.8%;  Pred. No. 4.7e-145;
  Matches  357;  Conservative   13;  Mismatches   23;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||:|| |||||||||| |||||||||   || || | ||||||| |||||||
Db        814 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFTTRAEADLHVWIPPLNARGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| |||||||||||||:|| ||||||||||:|||||||| ||||||| |||||
Db        874 ILLMCAVHPELIFDITKLLIAILGPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||| |||||:|:|||||:||  |||||||||||||:||||| ||||||||||||
Db        934 VQMAFMKLGALTGTYIYNHLTPLRDWPRAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||:|||||||: ||:| ||||||||||||||||||||||||||||||
Db        994 GDIILGLPVSARRGKEILLGPADSLEGRGLRLLAPITAYSQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||:||||||||| ||||||||||||||||||||||
Db       1054 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAAPKGPITQMYTNVDQDLVGWPK 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1114 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPF 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
Db       1174 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 1206


RESULT 5
POLG_HCV1
ID   POLG_HCV1      STANDARD;      PRT;  3011 AA.
AC   P26664;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate 1) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11104;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91172826; PubMed=1848704;
RA   Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA   Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
RA   Bradley D.W., Kuo G., Houghton M.;
RT   "Genetic organization and diversity of the hepatitis C virus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M62321; AAA45676.1; -.
DR   PIR; A39166; GNWVC3.
DR   PDB; 1A1V; 16-FEB-99.
DR   PDB; 1HEI; 25-NOV-98.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT   CHAIN       730   1006       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
```

```
FT   CHAIN       1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN       1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN       1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN       2014   3011       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM     347    369       POTENTIAL.
FT   ACT_SITE    1083   1083       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1107   1107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1165   1165       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND     1230   1237       ATP (POTENTIAL).
FT   SITE        1316   1319       DECH BOX.
FT   CARBOHYD     196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     476    476       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2041   2041       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2077   2077       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2240   2240       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2364   2364       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2789   2789       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3011 AA;  327197 MW;  65F8C9447FCE5AF9 CRC64;

  Query Match             85.9%;  Score 1764;  DB 1;  Length 3011;
  Best Local Similarity   81.9%;  Pred. No. 5.6e-135;
  Matches  322;  Conservative   34;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              :|||||| | :||  |||||||  ::  :|||| :|||| |  ||||||||||||||:
Db        814 VAASCGGVVLVGLMALTLSPYYKRYISWCLWWLQYFLTRVEAQLHVWIPPLNVRGGRDAV 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| ||||| |:|||||||||:|||| :||| : ||||||| |||:|:| | ||  ||||
Db        874 ILLMCAVHPTLVFDITKLLLAVFGPLWILQASLLKVPYFVRVQGLLRFCALARKMIGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||  :|| |||||||:|||||:|||| ||||||||||||:|| || |:||||||||||
Db        934 VQMVIIKLGALTGTYVYNHLTPLRDWAHNGLRDLAVAVEPVVFSQMETKLITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              ||||:||||||||||||||    :||||||||||:||||||||||||||||||||||||
Db        994 GDIINGLPVSARRGREILLGPADGMVSKGWRLLAPITAYAQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              |||||||:|||| |:|||||:|||||||:||||::|:| ||||: ||||||||||||| |
Db       1054 QVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLVGWPA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              | |:||:||||||||||||||||||||||||||||||||||||:|||||||||||||||:
Db       1114 PQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPA 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||||||::|||||
Db       1174 GHAVGIFRAAVCTRGVAKAVDFIPVENLETTMR 1206


RESULT 6
POLG_HCVH
ID   POLG_HCVH      STANDARD;      PRT;  3011 AA.
AC   P27958;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate H) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11108;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92052256; PubMed=1658800;
RA   Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA   Prince A.M.;
RT   "Genomic structure of the human prototype strain H of hepatitis C
RT   virus: comparison with American and Japanese isolates.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX   MEDLINE=97331322; PubMed=9187654;
RA   Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT   "Structure of the hepatitis C virus RNA helicase domain.";
RL   Nat. Struct. Biol. 4:463-467(1997).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX   MEDLINE=98154321; PubMed=9493270;
RA   Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA   Murcko M.A., Lin C., Caron P.R.;
RT   "Hepatitis C virus NS3 RNA helicase domain with a bound
RT   oligonucleotide: the crystal structure provides insights into the mode
RT   of unwinding.";
RL   Structure 6:89-100(1998).
CC   -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC   -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC       NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC   -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC       ACTIVATION OF NS3.
CC   -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC   -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC       ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC       AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC   -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC       PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC   -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC   -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M67463; AAA45534.1; -.
DR   PIR; A36814; GNWVCH.
DR   PDB; 1HEI; 25-NOV-98.
DR   PDB; 1A1V; 16-FEB-99.
DR   PDB; 1A1R; 17-JUN-98.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   TRANSFAC; T04155; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
```

```
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    191       CAPSID PROTEIN C.
FT   CHAIN       192    383       ENVELOPE GLYCOPROTEIN E1.
FT   CHAIN       384    746       ENVELOPE GLYCOPROTEIN E2.
FT   CHAIN       747    809       PROTEIN P7.
FT   CHAIN       810   1026       NONSTRUCTURAL PROTEIN NS2.
FT   CHAIN      1027   1657       PROTEASE/HELICASE NS3.
FT   CHAIN      1658   1711       NONSTRUCTURAL PROTEIN NS4A.
FT   CHAIN      1712   1972       NONSTRUCTURAL PROTEIN NS4B.
FT   CHAIN      1973   2420       NONSTRUCTURAL PROTEIN NS5A.
FT   CHAIN      2421   3011       NONSTRUCTURAL PROTEIN NS5B.
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
FT   SITE       1316   1319       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    476    476       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   STRAND     1224   1226
FT   TURN       1232   1233
FT   TURN       1236   1238
FT   HELIX      1239   1246
FT   TURN       1247   1248
FT   STRAND     1251   1255
FT   HELIX      1258   1271
FT   TURN       1272   1272
FT   STRAND     1277   1280
FT   TURN       1281   1282
FT   STRAND     1283   1285
FT   STRAND     1291   1295
FT   HELIX      1296   1301
FT   TURN       1302   1303
FT   STRAND     1312   1316
FT   TURN       1317   1319
FT   HELIX      1323   1335
FT   TURN       1336   1340
FT   STRAND     1343   1347
FT   TURN       1352   1353
FT   TURN       1360   1361
FT   STRAND     1362   1366
FT   STRAND     1368   1368
FT   STRAND     1373   1375
FT   TURN       1376   1377
FT   STRAND     1378   1380
FT   HELIX      1382   1385
FT   STRAND     1389   1393
FT   HELIX      1397   1409
FT   TURN       1410   1411
FT   STRAND     1414   1417
FT   TURN       1419   1420
FT   STRAND     1432   1436
FT   TURN       1438   1439
FT   STRAND     1450   1453
FT   STRAND     1456   1463
FT   STRAND     1471   1478
FT   STRAND     1480   1480
FT   HELIX      1481   1488
FT   TURN       1489   1490
FT   STRAND     1497   1501
FT   STRAND     1507   1507
FT   STRAND     1511   1511
FT   HELIX      1514   1527
FT   HELIX      1532   1544
FT   STRAND     1550   1550
FT   HELIX      1555   1564
FT   HELIX      1570   1578
FT   TURN       1579   1580
FT   HELIX      1584   1597
FT   TURN       1598   1598
FT   HELIX      1606   1611
FT   TURN       1614   1618
FT   STRAND     1622   1623
FT   STRAND     1627   1627
FT   STRAND     1635   1636
FT   HELIX      1640   1652
SQ   SEQUENCE   3011 AA;  327142 MW;  772CBB29CCD94753 CRC64;

  Query Match             85.3%;  Score 1752;  DB 1;  Length 3011;
  Best Local Similarity   81.4%;  Pred. No. 5.3e-134;
  Matches  320;  Conservative   36;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              :|||||| | :||  |||||||  ::  :|||||:|||||| | ||:||||||||||:
Db        814 VAASCGGVVLVGLMALTLSPYYKRYISWCMWWLQYFLTRVEAQLHVWVPPLNVRGGRDAV 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||| ||| |:|||||||||||||| :||| : ||||||| |||:| | | || |||||
Db        874 ILLTCVVHPALVFDITKLLLAIFGPLWILQASLLKVPYFVRVQGLLRICALARKIAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| :|| ||||| ||:|| ||:|||| |||||||||||:|| || |:||||||||||
Db        934 VQMAIIKLGALTGTCVYNHLAPLRDWAHNGLRDLAVAVEPVVFSRMETKLITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              ||||:|||||||||:|||||||     :|||||||||||:|||||||||||||||||||
Db        994 GDIINGLPVSARRGQEILLGPADGMVSKGWRLLAPITAYAQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              |||||||:||||||:|||||:|||||||:||||::|:| ||||: | ||||||||||| |
Db       1054 QVEGEVQIVSTATQTFLATCINGVCWTVYHGAGTRTIASPKGPVIQTYTNVDQDLVGWPA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
```

```
                 | |:||:|||||||||||||||||||||||||||||||||||:|||||||||||||||:
Db       1114 PQGSRSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPT 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
                 |||||:|||||||||||||||||::|||||
Db       1174 GHAVGLFRAAVCTRGVAKAVDFIPVENLETTMR 1206


RESULT 7
POLG_HCVJ6
ID   POLG_HCVJ6     STANDARD;      PRT;  3033 AA.
AC   P26660;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate HC-J6) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11113;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92044440; PubMed=1658196;
RA   Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
RA   Machida A., Miyakawa Y., Mayumi M.;
RT   "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
RT   from a human carrier: comparison with reported isolates for conserved
RT   and divergent regions.";
RL   J. Gen. Virol. 72:2697-2704(1991).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D00944; BAA00792.1; -.
DR   PIR; JQ1303; JQ1303.
DR   HSSP; P27958; 1HEI.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    733       NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT   CHAIN       734   1010       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1011   1619       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1620   1866       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1867   2017       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2018   3033       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1087   1087       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1111   1111       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1169   1169       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1234   1241       ATP (POTENTIAL).
FT   SITE       1320   1323       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    477    477       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    534    534       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    542    542       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    558    558       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    578    578       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    627    627       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    649    649       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1091   1091       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2038   2038       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2811   2811       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3033 AA;  329165 MW;  F957F5C1A273BE9E CRC64;

  Query Match           68.3%;  Score 1403;  DB 1;  Length 3033;
  Best Local Similarity   64.9%;  Pred. No. 1.3e-105;
  Matches  252;  Conservative   58;  Mismatches   78;  Indels    0;  Gaps    0;

Qy          6 GGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTC 65
              | |: : : | ||:| || ||:| :||| ||:|  || :| | ||: ||||| ||
Db        823 GAALLVLITLFTLTPGYKTLLSRFLWWLCYLLTLAEAMVQEWAPPMQVRGGRDGIIWAVA 882

Qy         66 AVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAF 125
                 | ::||||| |||: ||  :|:  :|:|||||||  |:| | :||  ||| ||||
Db        883 IFCPGVVFDITKWLLAVLGPAYLLKGALTRVPYFVRAHALLRMCTMVRHLAGGRYVQMVL 942

Qy        126 MKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIIS 185
             : |  ||||:|||||: |||  |||||||||||:||| || |:| |||:|||||||:
Db        943 LALGRWTGTYIYDHLTPMSDWAANGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILH 1002
```

```
Qy        186 GLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGE 245
              ||||||| |||:|||||| :  :|| |||||||:||||||| |: |:|||||  :  ||
Db       1003 GLPVSARLGREVLLGPADGYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGE 1062

Qy        246 VQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGAR 305
              :||:|| ||||| | ::|| |||:||||:||||| :||:||||:: : ||||| :||| :
Db       1063 IQVLSTVTQSFLGTTISGVLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTK 1122

Qy        306 SMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVG 365
              |: |||||: |||||||:||||| ||||| ||:||||||:| |||||||:||| |||||
Db       1123 SLEPCTCGAVDLYLVTRNADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVG 1182

Qy        366 IFRAAVCTRGVAKAVDFIPVESMETTMR 393
              :||||||:|||||::|||||:::   |
Db       1183 VFRAAVCSRGVAKSIDFIPVETLDIVTR 1210
```

```
RESULT 8
POLG_HCVJ8
ID   POLG_HCVJ8     STANDARD;      PRT;  3033 AA.
AC   P26661;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate HC-J8) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11115;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92230232; PubMed=1314459;
RA   Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
RA   Fukuda S., Tsuda F., Mishiro S.;
RT   "Full-length sequence of a hepatitis C virus genome having poor
RT   homology to reported isolates: comparative study of four distinct
RT   genotypes.";
RL   Virology 188:331-341(1992).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       (RNA)(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10988; BAA01761.1; -.
DR   PIR; A40250; GNWVJ8.
DR   HSSP; P27958; 1HEI.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    733       NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT   CHAIN       734   1010       NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1011   1619       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1620   1866       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1867   2017       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2018   3033       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1087   1087       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1111   1111       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1169   1169       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1234   1241       ATP (POTENTIAL).
FT   SITE       1320   1323       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    233    233       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    299    299       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    477    477       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    534    534       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    542    542       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    558    558       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    578    578       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    627    627       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    649    649       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1091   1091       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2038   2038       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2359   2359       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2811   2811       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3033 AA;  330177 MW;  1A173E7E3381FD1A CRC64;
```

```
  Query Match             68.2%;  Score 1401;  DB 1;  Length 3033;
  Best Local Similarity   63.4%;  Pred. No. 1.9e-105;
  Matches  246;  Conservative   67;  Mismatches   75;  Indels    0;  Gaps    0;

Qy          6 GGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTC 65
              | |: : ::: ||:| ||:||:| :||| |::   || :| |:||| |||||| || :
Db        823 GLAILVIISIFTLTPAYKILLSRSVWWLSYMLVLAEAQIQQWVPPLEVRGGRDGIIWVAV 882
```

```
Qy          66 AVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAF 125
               :|| |:|::|| |||| ||  :|:| : ::|||||  |:| | ||:  ||  |:||
Db         883 ILHPRLVFEVTKWLLAILGPAYLLKASLLRIPYFVRAHALLRVCTLVKHLAGARYIQMLL 942

Qy         126 MKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIIS 185
               : :   ||||:||||:||  ||  ||||||:|||||:|| || |:| |||:| |||||:
Db         943 ITIGRWTGTYIYDHLSPLSTWAAQGLRDLAIAVEPVVFSPMEKKVIVWGAETVACGDILH 1002

Qy         186 GLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGE 245
               ||||||| |||:|||||| :  :||:||||||||:||||||| |: |||||||:  |:
Db        1003 GLPVSARLGREVLLGPADGYTSKGWKLLAPITAYTQQTRGLLGAIVVSLTGRDKNEQAGQ 1062

Qy         246 VQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGAR 305
               |||:|: ||:|| | ::|| |||:||||:|||||||||:|||||: : ||||| :||| :
Db        1063 VQVLSSVTQTFLGTSISGVLWTVYHGAGNKTLAGPKGPVTQMYTSAEGDLVGWPSPPGTK 1122

Qy         306 SMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVG 365
               |: ||||| ||||||:|||||||: | ||:||||||:| |||||||:||  ||||
Db        1123 SLDPCTCGAVDLYLVTRNADVIPVRRKDDRRGALLSPRPLSTLKGSSGGPVLCSRGHAVG 1182

Qy         366 IFRAAVCTRGVAKAVDFIPVESMETTMR 393
               :|||||| |||||::|||||||::   |
Db        1183 LFRAAVCARGVAKSIDFIPVESLDVATR 1210
```

```
RESULT 9
VST2_HEVBU
ID   VST2_HEVBU     STANDARD;      PRT;   660 AA.
AC   P29326;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   Structural protein 2 precursor (ORF2).
OS   Hepatitis E virus (strain Burma) (HEV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage;
OC   Hepatitis E-like viruses.
OX   NCBI_TaxID=31767;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92024067; PubMed=1926770;
RA   Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,
RA   Fry K.E., Reyes G.R.;
RT   "Hepatitis E virus (HEV): molecular cloning and sequencing of the
RT   full-length viral genome.";
RL   Virology 185:120-131(1991).
CC   -!- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
CC       THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
CC       BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M73218; AAA45736.1; -.
DR   PIR; C40778; VHWWH2.
DR   InterPro; IPR004261; SP2.
DR   InterPro; IPR008975; Viral_cap_coat.
DR   Pfam; PF03014; SP2; 1.
KW   Signal.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    660       STRUCTURAL PROTEIN 2.
SQ   SEQUENCE   660 AA;  70978 MW;  5832A013CCC4A61C CRC64;

  Query Match             5.0%;  Score 102.5;  DB 1;  Length 660;
  Best Local Similarity   19.8%;  Pred. No. 1.1;
  Matches   82;  Conservative   52;  Mismatches  130;  Indels  151;  Gaps   19;

Qy          76 TKLLL--AIFGPLMVLQAG------ITKVPYFVR---AQGLIRACMLVRKAAGGHYVQMA 124
```

```
               | |:|  |   ||: || |       |:   : :   |:  ||   ||  | ||: : ::
Db         151 TNLVLYAAPLSPLLPLQDGTNTHIMATEASNYAQYRVARATIRYRPLVPNAVGGYAISIS 210

Qy         125 FMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDII 184
               |      | | |       |       |: : |:| | |:: :
Db         211 FWPQTTTTPTSV--------DMNSITSTDVRILVQPGIASELVI---------------- 246

Qy         185 SGLPVSARRGREILLGPAD--NFEGQGWRLLAPI-TAYSQQTRGLL-----GCIITSLTG 236
                           |::  ::  |||| :    |  : | ||:      | :: | |
Db         247 ----------------PSERLHYRNQGWRSVETSGVAEEEATSGLVMLCIHGSLVNSYTN 290

Qy         237 -------------------------------------RDKNQVEGEVQVVSTATQSFL 257
                                                   | :   :|  :: :||   |:
Db         291 TPYTGALGLLDFALELEFRNLTPGNTNTRVSRYSSTARHRLRRGADGTAELTTTAATRFM 350

Qy         258 A----TCVNGV-------CWTVFH--------------GAG-----SKTLAGPKG-PIT 285
               |  |||          |:|:                ||      |: :    | |
Db         351 KDLYFTSTNGVGEIGRGIALTLFNLADTLLGGLPTELISSAGGQLFYSRPVVSANGEPTV 410

Qy         286 QMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLV---TRHADVIPVRRRGDSRG-SLLS 341
               ::||:|:      |   |       | | : :     :|    |       ||  |:|
Db         411 KLYTSVENA----QQDKGIAIPHDIDLGESRVVIQDYDNQHEQDRPTPSPAPSRPFSVLR 466

Qy         342 PRPVSYLK-----------GSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPV 385
               | :|           ||| ||:      :| :   |   : ||:::|: |
Db         467 ANDVLWLSLTAAEYDQSTYGSSTGPVYV--SDSVTLVNVATGAQAVARSLDWTKV 519
```

```
RESULT 10
VST2_HEVPA
ID   VST2_HEVPA     STANDARD;      PRT;   660 AA.
AC   P33426;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   Structural protein 2 precursor (ORF2).
OS   Hepatitis E virus (strain Pakistan) (HEV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage;
OC   Hepatitis E-like viruses.
OX   NCBI_TaxID=33774;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92115700; PubMed=1731327;
RA   Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,
RA   Malik I.A., Iqbal M., Purcell R.H.;
RT   "Characterization of a prototype strain of hepatitis E virus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).
CC   -!- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
CC       THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
CC       BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M80581; AAA45727.1; -.
DR   InterPro; IPR004261; SP2.
DR   InterPro; IPR008975; Viral_cap_coat.
DR   Pfam; PF03014; SP2; 1.
KW   Signal.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   CHAIN        23    660       STRUCTURAL PROTEIN 2.
SQ   SEQUENCE   660 AA;  70980 MW;  8085BC53CFB46FD3 CRC64;

  Query Match             5.0%;  Score 102.5;  DB 1;  Length 660;
  Best Local Similarity   19.8%;  Pred. No. 1.1;
  Matches   82;  Conservative   54;  Mismatches  128;  Indels  151;  Gaps   20;

Qy          76 TKLLL--AIFGPLMVLQAG------ITKVPYFVR---AQGLIRACMLVRKAAGGHYVQMA 124
```

```
Db        151 TNLVLYAAPLSPLLPLQDGTNTHIMATEASNYAQYRVARATIRYRPLVPNAVGGYAISIS 210

Qy        125 FMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDII 184

Db        211 FWPQTTTTPTSV--------DMNSITSTDVRILVQPGIASELVI---------------- 246

Qy        185 SGLPVSARRGREILLGPAD--NFEGQGWRLL--------------------APITAYSQ 221

Db        247 ----------------PSERLHYRNQGWRSVETSGVAEEEATSGLVMLCIHGSPVNSYTN 290

Qy        222 QT-RGLLGCI-------ITSLTGRDKNQ----------------VEGEVQVVSTATQSFL 257

Db        291 TPYTGALGLLDFALELEFRNLTPGNTNTRVSRYSSTARHRLRRGADGTAELTTTAATRFM 350

Qy        258 A----TCVNGV-------CWTVFH----------------GAG-----SKTLAGPKG-PIT 285

Db        351 KDLYFTSTNGVGEIGRGIALTLFNLADTLLGGLPTELISSAGGQLFYSRPVVSANGEPTV 410

Qy        286 QMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLV---TRHADVIPVRRRGDSRG-SLLS 341

Db        411 KLYTSVENA----QQDKGIAIPHDIDLGESRVVIQDYDNQHEQDRPTPSPAPSRPFSVLR 466

Qy        342 PRPVSYLK-----------GSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPV 385

Db        467 ANDVLWLSLTAAEYDQSTYGSSTGPVYV--SDSVTLVNVATGAQAVARSLDWTKV 519
```

```
RESULT 11
POLG_MVEV
ID   POLG_MVEV      STANDARD;      PRT;  1780 AA.
AC   P05769;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
DE   protein (Envelope protein M); Major envelope protein E; Nonstructural
DE   proteins NS1, NS2A, and NS2B; Protease/helicase (EC 3.4.21.98) (NS3)]
DE   (Fragment).
OS   Murray valley encephalitis virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Flavivirus.
OX   NCBI_TaxID=11079;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86200215; PubMed=3009829;
RA   Dalgarno L., Trent D.W., Strauss J.H., Rice C.M.;
RT   "Partial nucleotide sequence of the Murray Valley encephalitis virus
RT   genome. Comparison of the encoded polypeptides with yellow fever
RT   virus structural and non-structural proteins.";
RL   J. Mol. Biol. 187:309-323(1986).
RN   [2]
RP   CARBOHYDRATES-LINKAGE SITES OF NS1, AND DISULFIDE BONDS.
RX   MEDLINE=21405829; PubMed=11514736;
RA   Blitvich B.J., Scanlon D., Shiell B.J., Mackenzie J.S., Pham K.,
RA   Hall R.A.;
RT   "Determination of the intramolecular disulfide bond arrangement and
RT   biochemical identification of the glycosylation sites of the
RT   nonstructural protein NS1 of Murray Valley encephalitis virus.";
RL   J. Gen. Virol. 82:2251-2256(2001).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X03467; CAA27184.1; -.
DR   PIR; A24635; GNWVMV.
DR   HSSP; P14336; 1SVB.
DR   MEROPS; S07.001; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001122; Flavi_capsidC.
DR   InterPro; IPR000336; Flavi_glycoprotE.
DR   InterPro; IPR000069; Flavi_M.
DR   InterPro; IPR001157; Flavi_NS1.
DR   InterPro; IPR000752; Flavi_NS2A.
DR   InterPro; IPR000487; Flavi_NS2B.
DR   InterPro; IPR002535; Flavi_propep.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001850; Peptidase_S7.
DR   Pfam; PF01003; Flavi_capsid; 1.
DR   Pfam; PF02832; Flavi_glycop_C; 1.
DR   Pfam; PF00869; Flavi_glycoprot; 1.
DR   Pfam; PF00949; Flavi_helicase; 1.
DR   Pfam; PF01004; Flavi_M; 1.
DR   Pfam; PF00948; Flavi_NS1; 1.
DR   Pfam; PF01005; Flavi_NS2A; 1.
DR   Pfam; PF01002; Flavi_NS2B; 1.
DR   Pfam; PF01570; Flavi_propep; 1.
DR   ProDom; PD001556; Flavi_glycoprotE; 1.
DR   ProDom; PD001496; Flavi_NS1; 1.
KW   Polyprotein; Glycoprotein; Core protein; Coat protein;
KW   Envelope protein; Hydrolase; Helicase; ATP-binding; Transmembrane;
KW   Nonstructural protein.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    121       CAPSID PROTEIN C.
FT   PROPEP      122    217
FT   CHAIN       218    292       ENVELOPE GLYCOPROTEIN M.
FT   CHAIN       293    793       MAJOR ENVELOPE PROTEIN E.
FT   CHAIN       794   1207       NONSTRUCTURAL PROTEIN NS1.
FT   CHAIN      1208   1372       NONSTRUCTURAL PROTEIN NS2A.
FT   CHAIN      1373   1503       NONSTRUCTURAL PROTEIN NS2B.
FT   CHAIN      1504  >1780       PROTEASE/HELICASE (NS3).
FT   TRANSMEM     44     60       POTENTIAL.
FT   TRANSMEM    112    128       POTENTIAL.
FT   TRANSMEM    278    294       POTENTIAL.
FT   TRANSMEM    773    789       POTENTIAL.
FT   TRANSMEM   1178   1194       POTENTIAL.
FT   TRANSMEM   1219   1235       POTENTIAL.
FT   TRANSMEM   1250   1266       POTENTIAL.
FT   TRANSMEM   1312   1328       POTENTIAL.
FT   TRANSMEM   1378   1394       POTENTIAL.
FT   TRANSMEM   1401   1417       POTENTIAL.
FT   TRANSMEM   1476   1492       POTENTIAL.
FT   DISULFID    295    322       BY SIMILARITY.
FT   DISULFID    352    408       BY SIMILARITY.
FT   DISULFID    366    397       BY SIMILARITY.
FT   DISULFID    384    413       BY SIMILARITY.
FT   DISULFID    482    580       BY SIMILARITY.
FT   DISULFID    597    628       BY SIMILARITY.
FT   DISULFID    797    808
FT   DISULFID    848    936
FT   DISULFID    972   1016
FT   CARBOHYD     73     73       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    140    140       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    446    446       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    923    923       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    968    968       N-LINKED (GLCNAC. . .).
FT   CARBOHYD   1000   1000       N-LINKED (GLCNAC. . .) (HIGH MANNOSE).
FT   NON_TER    1780   1780
SQ   SEQUENCE   1780 AA;  194866 MW;  0D6AA7FF0FB706DE CRC64;
```

```
  Query Match             4.9%;  Score 101.5;  DB 1;  Length 1780;
  Best Local Similarity   19.8%;  Pred. No. 4.4;
  Matches   81;  Conservative   49;  Mismatches  136;  Indels  143;  Gaps   16;

Qy          7 GAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCA 66
              ||| ||||| :   :   ::|              | | :  |  | : |  |  :||
Db       1340 GAVLIGLALTSTGYFSPTIMA-------------AGLMICNP--NKKRGWPATEVLTA- 1382

Qy         67 VHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFM 126
                      |: || | |  |       ||:       |   |||     |    |
Db       1383 ---------VGLMFAIVGGLAELDIDSMSVPF------TIAGLMLVSYVISGKATDMWLE 1427

Qy        127 KL--------AALTGT--------------YVYDHLTPLQDWAHAGLRDLAVAVEPVIF 163
               :         ||:|||                 : |   | : |    ||   ::|
Db       1428 RAADVSWEAGAAITGTSERLDVQLDDDGDFHLLNDPGVPWKIWV---LRMTCLSV----- 1479

Qy        164 SDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEGQG------WRLLAPIT 217
                  |  |    :| | ::     :||       |:     :|       :|::|
Db       1480 ----AAITPWAILPSAFGYWLT--LKYTKRGGVFWDTPSPKVYPKGDTTPGVYRIMA--- 1530

Qy        218 AYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTL 277
                ||:||                          |: :     ||  |::|     :
Db       1531 ------RGILG----------------------RYQAGVGVMHEGVFHTLWHTTRGAAI 1561

Qy        278 AGPKGPITQMYTNVDQDLVGWQAP------------------PGARSMTPCTCGSSDLY 318
                 :| :|  : || :| | :  |                    ||  ::   |
Db       1562 MSGEGRLTPYWGNVKEDRVTYGGPWKLDQKWNGVDDVQMIVVEPGKPAINVQT---KPGI 1618

Qy        319 LVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIF 367
                | | ::          |::    |:    |:|| |::  :|  :|::
Db       1619 FKTAHGEI----------GAVSLDYPI----GTSGSPIVNSNGEIIGLY 1653


RESULT 12
SR5C_ARATH
ID   SR5C_ARATH     STANDARD;      PRT;   564 AA.
AC   P37107; O82570;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Signal recognition particle 54 kDa protein, chloroplast precursor
DE   (SRP54) (54 chloroplast protein) (54CP) (FFC).
GN   FFC OR AT5G03940 OR F8F6_150.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=94012817; PubMed=8408079;
RA   Franklin A.E., Hoffman N.E.;
RT   "Characterization of a chloroplast homologue of the 54-kDa subunit of
RT   the signal recognition particle.";
RL   J. Biol. Chem. 268:22175-22180(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Amin P., Sy D., Pilgrim M., Parry D.H., Hoffman N.E.;
RT   "Isolation of two Arabidopsis mutants in the nuclear gene ffc,
RT   encoding the 54 kDa subunit of chloroplast signal recognition
RT   particle.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016721; PubMed=11130714;
RA   Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
RA   Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA   Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA   Nakazaki N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,
RA   Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
RA   Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RA   Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RA   Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
RA   Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
RA   Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
RA   Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strowmatt C.,
RA   Wagner-McPherson C., Wollam A., Yoakum M., Bell M., Dedhia N.,
RA   Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
RA   Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.A.,
RA   Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RA   Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
RA   Entian K.-D., Terryn N., Hartley N., Bent E., Johnson S.,
RA   Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,
RA   Ramsperger U., Wedler H., Balke K., Wedler E., Peters S.,
RA   van Staveren M., Dirkse W., Mooijman P., Klein Lankhorst R.,
RA   Weitzenegger T., Bothe G., Rose M., Hauf J., Berneiser S., Hempel S.,
RA   Feldpausch M., Lamberth S., Villarroel R., Gielen J., Ardiles W.,
RA   Bents O., Lemcke K., Kolesov G., Mayer K., Rudd S., Schoof H.,
RA   Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Fransz P.;
RT   "Sequence and analysis of chromosome 5 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:823-826(2000).
CC   -!- FUNCTION: May target chloroplast proteins to either the thylakoid
CC       or envelope membranes.
CC   -!- SUBCELLULAR LOCATION: Chloroplast stroma.
CC   -!- TISSUE SPECIFICITY: Most abundant in green shoot tissue and
CC       lower levels seen in the roots and etiolated buds.
CC   -!- SIMILARITY: Belongs to the GTP-binding SRP family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z21970; CAA79981.1; -.
DR   EMBL; AF092168; AAC64139.1; -.
DR   EMBL; AL162873; CAB85514.1; -.
DR   PIR; S36637; S36637.
DR   HSSP; O07347; 1FFH.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR000897; SRP54.
DR   InterPro; IPR004125; SRP54_SPB.
DR   InterPro; IPR004780; SRP_sub.
DR   Pfam; PF00448; SRP54; 1.
DR   Pfam; PF02881; SRP54_N; 1.
DR   Pfam; PF02978; SRP_SPB; 1.
DR   ProDom; PD000819; SRP54; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR00959; ffh; 1.
DR   PROSITE; PS00300; SRP54; 1.
KW   Signal recognition particle; GTP-binding; RNA-binding; Chloroplast;
KW   Transit peptide.
FT   TRANSIT        1     75       CHLOROPLAST.
FT   CHAIN         76    564       SIGNAL RECOGNITION PARTICLE 54 kDa
FT                                 PROTEIN.
FT   DOMAIN        76    370       G-DOMAIN.
FT   DOMAIN       371    564       M-DOMAIN.
FT   NP_BIND      183    190       GTP (BY SIMILARITY).
FT   NP_BIND      265    269       GTP (BY SIMILARITY).
FT   NP_BIND      323    326       GTP (BY SIMILARITY).
FT   CONFLICT      76     76       E -> V (IN REF. 2).
SQ   SEQUENCE   564 AA;  61232 MW;  423F7285FB9063E4 CRC64;

  Query Match             4.9%;  Score 101;  DB 1;  Length 564;
  Best Local Similarity   26.1%;  Pred. No. 1.2;
  Matches   54;  Conservative   37;  Mismatches   74;  Indels   42;  Gaps   11;

Qy         67 VHPELI----FDITKLLLAIFGPLMVLQAGI---------TKVPYFVRAQGLIRACMLVR 113
              || ||:     ::::|   |  || ::| ||:           |: ::: ||  ::|||:
Db        154 VHDELVKLMGGEVSELQFAKSGPTVILLAGLQGVGKTTVCAKLACYLKKQG--KSCMLI- 210
```

```
Qy        114 KAAGGHYVQMAFMKLAAL---TGTYVYDHLTPLQ--DWAHAGLRDLAVAVEPVIFSDMEV 168
              ||  |   |  :|  |     |  ||   | ::  | |  ||::          ::::|
Db        211 --AGDVYRPAAIDQLVILGEQVGVPVYTAGTDVKPADIAKQGLKEAKK-------NNVDV 261

Qy        169 KIITWGADTAACGDIISGLPVSARRGREIL-----LGPADNFEGQGWRLLAPITAYSQQT 223
              |:    |||     |  |:    :  :: |     |   |    ||     | :| :: :
Db        262 VIM----DTAGRLQIDKGMMDELKDVKKFLNPTEVLLVVDAMTGQ--EAAALVTTFNVEI 315

Qy        224 RGLLGCIITSLTGRDKNQVEGEVQVVS 250
              |: | |:| | |  :      |: ||
Db        316 -GITGAILTKLDGDSRGGAALSVKEVS 341


RESULT 13
DPO2_MOUSE
ID   DPO2_MOUSE      STANDARD;      PRT;   600 AA.
AC   P33611;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   DNA polymerase alpha 70 kDa subunit (DNA polymerase subunit B).
GN   POLA2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 84-102; 269-285 AND 394-403.
RX   MEDLINE=93216788; PubMed=8463324;
RA   Miyazawa H., Izumi M., Tada S., Takada R., Masutani M., Ui M.,
RA   Hanaoka F.;
RT   "Molecular cloning of the cDNAs for the four subunits of mouse DNA
RT   polymerase alpha-primase complex and their gene expression during
RT   cell proliferation and the cell cycle.";
RL   J. Biol. Chem. 268:8111-8122(1993).
CC   -!- FUNCTION: May play an essential role at the early stage of
CC       chromosomal DNA replication by coupling the polymerase
CC       alpha/primase complex to the cellular replication machinery (By
CC       similarity).
CC   -!- SUBUNIT: DNA polymerase alpha-primase is a four subunit enzyme
CC       (subunits A, B, C and D), which is assembled throughout the cell
CC       cycle. The largest subunit (subunit A) has DNA polymerase
CC       activity, the two smallest subunits (subunits C and D) have DNA
CC       primase activity. Subunit B binds to subunit A.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- PTM: PHOSPHORYLATED IN A CELL CYCLE-DEPENDENT MANNER, IN G2/M
CC       PHASE (BY SIMILARITY).
CC   -!- SIMILARITY: Belongs to the DNA polymerase alpha subunit B family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D13546; BAA02746.1; -.
DR   PIR; B46642; B46642.
DR   MGD; MGI:99690; Pola2.
DR   InterPro; IPR007200; DNA_pol_alpha_B.
DR   Pfam; PF04058; DNA_pol_alpha_B; 1.
KW   DNA replication; Nuclear protein; Phosphorylation.
FT   DOMAIN      101    107       POLY-GLU.
FT   DOMAIN      115    157       PRO/SER/THR-RICH (HYDROPHILIC).
SQ   SEQUENCE   600 AA;  66267 MW;  79F94BE6EF33FEBC CRC64;

  Query Match             4.9%;  Score 101;  DB 1;  Length 600;
  Best Local Similarity   24.8%;  Pred. No. 1.3;
  Matches   55;  Conservative   34;  Mismatches   71;  Indels   62;  Gaps   12;

Qy        105 LIRACMLVRKAAGGHYVQM-AFMKLAALT-------GTYVYDHL---------TPLQDWA 147
              |   |:| |:   |   :: ||   |  |        :: |: |         :  :|


Db         27 LAELCVLYRQTEDGMVSELIAFCTSAGKTCLTVDILNSFEYEVLNKKLSKAWHSASKDSG 86

Qy        148 HAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDI--ISGLP--------VSARRGREI 197
              ||| ||: |::: :| :: | : :      | : | :  :|  |         |:||  |:
Db         87 HAGTRDI-VSIQELIEAEEEEETLLSSYTTPSKGPLKRVSSTPETPLTKRSVAARSPRQ- 144

Qy        198 LLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFL 257
              || |:          :|     ||:       || | |      |||       : |
Db        145 LLSPSS---------FSPSATPSQK--------YTSRTNR------GEVVTTFGSAQ--- 178

Qy        258 ATCVNGVCWTVFHGAGSKTL--AGPKGPITQMYTNVDQDLVG 297
                    |: |:   |:|| :|   |   |:|  |  : | |:|
Db        179 -----GLSWSGRGGSGSVSLKVVGDPEPLTGSYKAMFQQLMG 215


RESULT 14
YA47_XYLFA
ID   YA47_XYLFA      STANDARD;      PRT;   444 AA.
AC   Q9PEI1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical zinc metalloprotease Xf1047 (EC 3.4.24.-).
GN   XF1047.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9a5c;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
CC   -!- COFACTOR: Zinc (Probable).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (By similarity).
CC   -!- SIMILARITY: Belongs to peptidase family M50B.
CC   -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ---------------------------------------------------------------------
DR   EMBL; AE003942; AAF83857.1; ALT_INIT.
DR   PIR; G82728; G82728.
DR   MEROPS; M50.004; -.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR004387; Pept_M50_Zn.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR008915; Peptidase_M50.
DR   Pfam; PF02163; Peptidase_M50; 1.
DR   SMART; SM00228; PDZ; 2.
DR   TIGRFAMs; TIGR00054; TIGR00054; 1.
DR   PROSITE; PS50106; PDZ; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hypothetical protein; Hydrolase; Metalloprotease; Zinc; Transmembrane;
KW   Inner membrane; Complete proteome.
FT   METAL        22     22       ZINC (CATALYTIC) (POTENTIAL).
FT   ACT_SITE     23     23       POTENTIAL.
FT   METAL        26     26       ZINC (CATALYTIC) (POTENTIAL).
FT   TRANSMEM     98    120       POTENTIAL.
FT   TRANSMEM    371    393       POTENTIAL.
FT   TRANSMEM    418    440       POTENTIAL.
FT   DOMAIN      192    276       PDZ.
SQ   SEQUENCE   444 AA;  48410 MW;  B10D1532DDBA4A34 CRC64;

  Query Match             4.6%;  Score 94;  DB 1;  Length 444;
  Best Local Similarity   20.4%;  Pred. No. 3.3;
  Matches   64;  Conservative   52;  Mismatches  117;  Indels   80;  Gaps   16;

Qy          8 AVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAV 67
              :|:  :|::   |   :||  |: |: ::| : :     |          ||   :     :
Db         93 SVWQRIAIVAAGPLANLLLCMLLLWVLFVIGKQDYSATV---------GRAEHLAAQAGI 143

Qy         68 HPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAG----GHYVQM 123
              ||        : || |          :|  :  |  |: |  : |: |      | | :
Db        144 HP------GDRITAIDG---------RQVTSWSEASMLLTAAAMDRQNAVLRVIGPYGER 188

Qy        124 A--FMKLAALTGTYVYDHLTPLQ--DWAHAGLRDLAVAVEPVIFSDMEVK-----IITWG 174
              :   ::|: |   :   |:| |   :|      :   :||   ::  :|     :   |
Db        189 SEHTLELSKLKQPFDERHVTALVGINWQFMLQPPIIAKIEPGSIAEGAIKPGDIVLAVDG 248

Qy        175 ADTAACGDIIS--------GLP--VSARRGREIL---LGPADNFEGQGWRL-----LAPI 216
                | :  |: :        | |  :  ||| | |   | |  : :| | |       |:
Db        249 QQTLSTEDLYNQIQKLGRDGHPGMIEIRRGEERLALELSPRKSAQGV-WLLGVKTNPGPV 307

Qy        217 TAY-SQQTRGLLGCIITSL--TGR----------------DKNQVEGEVQVVSTATQSF 256
                |: |||   |:|  :  ::  |||                : | : :   |
Db        308 PAFDSQQRYGVLAAVPLAIRETGRMTADSLGMMKRIITGQASAKNISGPISIAKIAN--- 364

Qy        257 LATCVNGVCWTVF 269
               |:  || | ::
Db        365 -ASAKRGVGWFIY 376


RESULT 15
CYAA_LEIDO
ID   CYAA_LEIDO     STANDARD;      PRT;  1380 AA.
AC   Q27675;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Receptor-type adenylate cyclase A (EC 4.6.1.1) (ATP pyrophosphate-
DE   lyase) (Adenylyl cyclase).
GN   RAC-A.
OS   Leishmania donovani.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX   NCBI_TaxID=5661;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1S Sudanese;
RX   MEDLINE=95340554; PubMed=7615561;
RA   Sanchez M.A., Zeoli D., Klamo E.M., Kavanaugh M.P., Landfear S.M.;
RT   "A family of putative receptor-adenylate cyclases from Leishmania
RT   donovani.";
RL   J. Biol. Chem. 270:17551-17558(1995).
CC   -!- FUNCTION: Could act as a receptor for a unknown ligand.
CC   -!- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC   -!- COFACTOR: Binds 1 magnesium ion per subunit (By similarity).
CC   -!- SUBCELLULAR LOCATION: Membrane-bound.
CC   -!- DEVELOPMENTAL STAGE: Expressed in the insect stage (promastigote)
CC       but not in the mammalian host stage of the parasite life cycle.
CC   -!- SIMILARITY: Belongs to the adenylyl cyclase class-3 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U17042; AAA74998.1; -.
DR   PIR; T18309; T18309.
DR   InterPro; IPR001054; G_cyclase.
DR   Pfam; PF00211; guanylate_cyc; 1.
DR   SMART; SM00044; CYCc; 1.
DR   PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
KW   Lyase; cAMP biosynthesis; Transmembrane; Receptor; Glycoprotein;
KW   Metal-binding; Magnesium.
FT   DOMAIN       1     34       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    35     55       POTENTIAL.
FT   DOMAIN      56    891       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   892    912       POTENTIAL.
FT   DOMAIN     913   1380       CYTOPLASMIC (POTENTIAL).
FT   METAL      938    938       MAGNESIUM (BY SIMILARITY).
FT   METAL      981    981       MAGNESIUM (BY SIMILARITY).
FT   CARBOHYD   422    422       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   478    478       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   497    497       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   567    567       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE  1380 AA;  151692 MW;  6B2D5F7D3C1107A0 CRC64;

  Query Match             4.6%;  Score 93.5;  DB 1;  Length 1380;
  Best Local Similarity   21.0%;  Pred. No. 15;
  Matches   93;  Conservative   51;  Mismatches  139;  Indels  159;  Gaps   24;

Qy         55 GGRDAIILLTCAVHP----ELIFDITKL----------LLAIFGPLM------------V 88
              |||   ||     ||    : ::|| ::          |||: || :             |
Db        101 GGRPIKIL-----HPDPDQDNLYDIAEVILHSLARQEKLLAVLGPYLDGRLTAALSNADV 155

Qy         89 LQAGITKVP--------------YFVRAQGLIRACMLVRKAAGGHYV------QMAFMK 127
              :|:|:  :              || ||: ::   :::       | |       ::|||:
Db        156 VQSGLMLIAPFTGSSGVRTWSDSVYFTRAEPMVELKVVLM-----HIVNRLRARRVAFMR 210

Qy        128 LAALTG--------TYVYDHLTPLQDWAHAGLRDLAVA-VEPVIFSDMEVK--------- 169
                 |||        ||| | || |       ||| ||    |   | :||
Db        211 ---LTGMHFGGEELTYVQDTLTSL-------LRDPAVLYTVPYSESSVEVDEEAFDAMAD 260

Qy        170 -----IITWGADTAACGDIISGLPVSARRGREILLG---------------PADNFEGQ 208
                  || | |        :  :    |     ::                  | : : |
Db        261 TNPQVIIVWAAPVQQVIYFLEKVLTDPRTSSAYVISCSMIQRVVFDVYKRLLSAGSIKPQ 320

Qy        209 GWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGE----VQVVSTATQSFLATCVNGV 264
              |:||  |                          : | ||    ::|:     :::    |
Db        321 DGRILASATT-------------------SPVSGEGLKYMEVLKAQMSNYIE---NSG 356

Qy        265 CWTVFHGAGSKTLAGPK----GPITQMYTNVDQDLVGWQAPPGARSMTPCTCGSSDLYLV 320
               :  :    |    | |    |:::  || ||:    :|| | |    :    |   |
Db        357 SFDYYPDDDSTETLGRKARSEAPLSRKYT-VDE---FFQAHPSIAKLMALGWLSGTLVQQ 412

Qy        321 TRHADVIPVRRRGDSRGSLLSPRPV---SYLKGSSGGPLLC-PSGHAVGIFRAAVCTRGV 376
              |     | |    |    | |    |: |  |||  | |    :|   :  | :|
Db        413 TLEQTDWIVNRSTYKAGLFNQNRFVIGGDYVLGDYGGP--CEPLAQFLG--ASCYCNQGG 468

Qy        377 AKAV---------DFIPVESME 389
               ::           | :|  | :
Db        469 HSSILTVLQNASWDIVPDSSFK 490
```

Search completed: May 6, 2004, 09:31:50
Job time : 9.20459 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:21:36 ; Search time 36.7566 Seconds
(without alignments)
3373.509 Million cell updates/sec

Title: US-10-650-585-11
Perfect score: 2053
Sequence: 1 MAASCGGAVFIGLALLTLSP.....RGVAKAVDFIPVESMETTMR 393

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1986	96.7	3010	12	Q9J3F9	Q9j3f9 hepatitis c
2	1981	96.5	3010	12	Q9J3H7	Q9j3h7 hepatitis c
3	1979	96.4	3010	12	Q9DTE6	Q9dte6 hepatitis c
4	1974	96.2	3010	12	Q9DTD6	Q9dtd6 hepatitis c
5	1973	96.1	3010	12	Q9DTE4	Q9dte4 hepatitis c
6	1973	96.1	3010	12	Q807P3	Q807p3 hepatitis c
7	1971	96.0	3010	12	Q9J3H3	Q9j3h3 hepatitis c
8	1970	96.0	3010	12	Q68788	Q68788 hepatitis c
9	1968	95.9	1186	12	Q81755	Q81755 hepatitis c
10	1968	95.9	2284	12	Q81817	Q81817 hepatitis c
11	1968	95.9	3010	12	P89966	P89966 hepatitis c
12	1968	95.9	3010	12	Q9J3G6	Q9j3g6 hepatitis c
13	1967	95.8	3010	12	Q99AU2	Q99au2 hepatitis c
14	1967	95.8	3010	12	Q9DTD7	Q9dtd7 hepatitis c
15	1967	95.8	3010	12	Q9QIX6	Q9qix6 hepatitis c
16	1967	95.8	3010	12	Q9QIX5	Q9qix5 hepatitis c
17	1966	95.8	3008	12	Q9J3F4	Q9j3f4 hepatitis c
18	1965	95.7	3010	12	P90191	P90191 hepatitis c
19	1964	95.7	3010	12	Q9J3H0	Q9j3h0 hepatitis c
20	1963	95.6	3010	12	Q9J3H9	Q9j3h9 hepatitis c
21	1963	95.6	3010	12	Q81760	Q81760 hepatitis c
22	1961	95.5	3010	12	Q9QIY3	Q9qiy3 hepatitis c
23	1961	95.5	3014	12	Q9DTE0	Q9dte0 hepatitis c
24	1960	95.5	3010	12	P88803	P88803 hepatitis c
25	1960	95.5	3010	12	Q9J3I0	Q9j3i0 hepatitis c
26	1960	95.5	3010	12	Q9J3H5	Q9j3h5 hepatitis c
27	1959	95.4	3010	12	Q9QIX8	Q9qix8 hepatitis c
28	1959	95.4	3010	12	Q9QIX7	Q9qix7 hepatitis c
29	1958	95.4	3010	12	Q9QIY5	Q9qiy5 hepatitis c
30	1958	95.4	3010	12	Q9DTE9	Q9dte9 hepatitis c
31	1958	95.4	3010	12	Q9J3H6	Q9j3h6 hepatitis c
32	1957	95.3	3010	12	Q9DTF0	Q9dtf0 hepatitis c
33	1955	95.2	3010	12	Q68826	Q68826 hepatitis c
34	1954	95.2	3015	12	Q9WPH5	Q9wph5 hepatitis c
35	1952	95.1	3010	12	Q9QIY4	Q9qiy4 hepatitis c
36	1951	95.0	3010	12	Q9J3I1	Q9j3i1 hepatitis c
37	1951	95.0	3010	12	Q9WMX2	Q9wmx2 hepatitis c
38	1951	95.0	3011	12	Q9DTE3	Q9dte3 hepatitis c
39	1950	95.0	3010	12	O09796	O09796 hepatitis c
40	1949	94.9	3010	12	Q9DTE5	Q9dte5 hepatitis c
41	1947	94.8	3010	12	Q9WIK8	Q9wik8 hepatitis c
42	1947	94.8	3010	12	Q9QIY6	Q9qiy6 hepatitis c
43	1946	94.8	1275	12	Q06642	Q06642 hepatitis c
44	1946	94.8	3010	12	Q9QIY7	Q9qiy7 hepatitis c
45	1946	94.8	3010	12	Q9QIY8	Q9qiy8 hepatitis c

ALIGNMENTS

RESULT 1
Q9J3F9
ID Q9J3F9 PRELIMINARY; PRT; 3010 AA.
AC Q9J3F9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD33;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
RT progression.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF207774; AAF65964.1; -.
DR PIR; A61196; A61196.
DR PIR; PQ0246; PQ0246.
DR PIR; PS0329; PS0329.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.

```
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327102 MW;  7162C9DB93E6E0C7 CRC64;

  Query Match           96.7%;  Score 1986;  DB 12;  Length 3010;
  Best Local Similarity  94.9%;  Pred. No. 5.4e-157;
  Matches  373;  Conservative   12;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|||||||||||||| ||||||||||||| |||| ||:||||||||||||
Db        814 MAASCGGAVFVGLALLTLSPYYKVFLARLIWWLQYLITRAEAHLHVWVPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||||||||||||||||||||||||||::|||||||||||||||| |||||
Db        874 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITRMPYFVRAQGLIRACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||| ||||||:|||||||:||||||||||||||||:||||| ||||||||||||
Db        934 VQMAFMKLGALTGTYIYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              ||||||||||||||:||||||||: ||:|||||||||||||||||||||||||||||||
Db        994 GDIISGLPVSARRGKEILLGPADSLEGRGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db       1054 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWPA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1114 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||:||||||||||
Db       1174 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 1206
```

```
RESULT 2
Q9J3H7
ID   Q9J3H7      PRELIMINARY;      PRT;  3010 AA.
AC   Q9J3H7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD15;
RA   Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT   "Characteristics of hepatitis C viral genome associated with disease
RT   progression.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF207756; AAF65946.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PQ0804; PQ0804.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
```

```
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327365 MW;  D8653F7317FFA106 CRC64;

  Query Match             96.5%;  Score 1981;  DB 12;  Length 3010;
  Best Local Similarity   94.7%;  Pred. No. 1.4e-156;
  Matches  372;  Conservative   11;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||||||||||||||||| |||||||||| ||| ||||||:||||||||||||
Db        814 MAASCGGAVFIGLALLTLSPYYKVFLARLIWWLQYFITRAEAHLQVWVPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              |:| ||||| |||||||||||| |||||||||::|||||||||||||||||||||||||
Db        874 IILACAVHPGLIFDITKLLLAILGPLMVLQAGITRMPYFVRAQGLIRACMLVRKAAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              :|||||||||||||:|||| ||| |||||||||||||||:||||| ||||||||||||
Db        934 IQMAFMKLAALTGTYIYDHLAPLQHWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||:||||||:||||||||||||||||||||||||||||||||||||
Db        994 GDIILGLPVSARRGREVLLGPADSFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              |||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||
Db       1054 QVEGEVQVVSTATQSFLATCINGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1114 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
Db       1174 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 1206


RESULT 3
Q9DTE6
ID   Q9DTE6      PRELIMINARY;      PRT;  3010 AA.
AC   Q9DTE6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCVT142;
RA   Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA   Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA   Mishiro S.;
RT   "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT   with hepatocellular carcinoma: the 'progression score' revisited.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AB049091; BAB18804.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW   Hydrolase; Nonstructural protein; Polyprotein;
KW   RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327042 MW;  3807DC6879684C95 CRC64;

  Query Match             96.4%;  Score 1979;  DB 12;  Length 3010;
  Best Local Similarity   94.4%;  Pred. No. 2.1e-156;
  Matches  371;  Conservative   11;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||| ||:|| |||||||||| ||:||||||| ||| |||||||:||||||||||||
Db        814 MAASCGGGVFVGLVLLTLSPYYKVFLAKLIWWLQYFITRAEAHLQVWVPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              |||||| |||||||||||||||||||||||||||||:|||||||||||||||||| |||||
Db        874 ILLTCAAHPELIFDITKLLLAIFGPLMVLQAGITRVPYFVRAQGLIRACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              ||||||||||||||||||||:|||| ||||||||||:||||| ||||||||||||||||
Db        934 VQMAFMKLAALTGTYVYDHLTPLRDWAHTGLRDLAVAVEPVVFSDMETKIITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||||||||||||||||||||: |||||||||||||||||||||||||||||||||||||
Db        994 GDIISGLPVSARRGREILLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db       1054 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1114 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              || |||||||||||||||||||:|||:||||||
Db       1174 GHVVGIFRAAVCTRGVAKAVDFVPVEAMETTMR 1206


RESULT 4
Q9DTD6
ID   Q9DTD6      PRELIMINARY;      PRT;  3010 AA.
AC   Q9DTD6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCVT221;
RA   Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA   Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA   Mishiro S.;
RT   "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT   with hepatocellular carcinoma: the 'progression score' revisited.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AB049101; BAB18814.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW   Hydrolase; Nonstructural protein; Polyprotein;
KW   RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327108 MW;  DE182D810EF78EE4 CRC64;

  Query Match           96.2%;  Score 1974;  DB 12;  Length 3010;
  Best Local Similarity   94.9%;  Pred. No. 5.4e-156;
  Matches  373;  Conservative    8;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| |||||||||| |||||||||| ||| |||| ||||||||||||||
Db        814 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFITRAEAHLHVWIPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||||||||||||:||||||| ||||| | |||||||||||||||||||| |||||
Db        874 ILLTCAVHPELIFDITKILLAIFGPPMVLQAVIIKVPYFVRAQGLIRACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              ||||||||||||||||||||||||||||||||||||||||||:||||| ||||||||||||
Db        934 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              ||||||||||||||||||||||:|||||||||||||||||||||||||||:|||||||||
Db        994 GDIISGLPVSARRGREILLGPADSFEGQGWRLLAPITAYSQQTRGLLGCIVTSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db       1054 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWPA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db       1114 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDTRGSLLSPRPVSYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              || ||:|||||||||||||||||||||||||||
Db       1174 GHIVGVFRAAVCTRGVAKAVDFIPVESMETTMR 1206


RESULT 5
Q9DTE4
ID   Q9DTE4      PRELIMINARY;      PRT;  3010 AA.
AC   Q9DTE4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCVT150;
RA   Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA   Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA   Mishiro S.;
RT   "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT   with hepatocellular carcinoma: the 'progression score' revisited.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
```

```
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AB049093; BAB18806.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PQ0804; PQ0804.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327324 MW;  3DE6CF249BD1151C CRC64;

  Query Match           96.1%;  Score 1973;  DB 12;  Length 3010;
  Best Local Similarity   94.4%;  Pred. No. 6.6e-156;
  Matches  371;  Conservative   11;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||:|||||||||||||| |||||||||||:|| |||||||:||||||||||||
Db        814 MAASCGGAVFVGLALLTLSPYYKVFLARLIWWLQYLLTRAEAHLQVWVPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| |||||||||||||||  ||||||:|||||||||||||||||||| |||||
Db        874 ILLMCAVHPELIFDITKLLLATLGPLMVLRAGITKVPYFVRAQGLIRACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||||||||||||||||||||||| ||||||||||:||||| ||||||||||||
Db        934 VQMAFMKLAALTGTYVYDHLTPLQDWAHASLRDLAVAVEPVVFSDMETKIITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||:||||||||: ||||||||||||||||||||||||||||||||||||
Db        994 GDIILGLPVSARRGKEILLLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              |||||||||||||||||||:|||||||:|||||||||||||||||||||||||||||||||
Db       1054 QVEGEVQVVSTATQSFLATCINGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:|||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       1114 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDGRGSLLSPRPVSYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
Db       1174 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 1206


RESULT 6
Q807P3
ID   Q807P3      PRELIMINARY;      PRT;  3010 AA.
AC   Q807P3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M1LE;
RX   MEDLINE=22047193; PubMed=12051758;
RA   Kishine H., Sugiyama K., Hijikata M., Kato N., Takahashi H., Noshi T.,
RA   Nio Y., Hosaka M., Miyanari Y., Shimotohno K.;
RT   "Subgenomic replicon derived from a cell line infected with the
RT   hepatitis C virus.";
RL   Biochem. Biophys. Res. Commun. 293:993-999(2002).
DR   EMBL; AB080299; BAC54896.1; -.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
```

```
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Polyprotein.
SQ   SEQUENCE   3010 AA;  327097 MW;  EE6418C7A723E686 CRC64;

  Query Match             96.1%;  Score 1973;  DB 12;  Length 3010;
  Best Local Similarity   94.7%;  Pred. No. 6.6e-156;
  Matches  372;  Conservative    9;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||||:||| |||||||||| |||||||||||||| |||||:|:||||:|||||||
Db        814 MAASCGGAVFVGLAFLTLSPYYKVFLARLIWWLQYLITRAEAHLQMWVPPLNIRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| |||||||||||||||||| ||||||||||:||||||||||||||||||||||||||
Db        874 ILLMCAVHPELIFDITKLLLAILGPLMVLQAGITRVPYFVRAQGLIRACMLVRKAAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||:|||||||||||||| ||| ||| |||||||||||:|||||| ||||||||||||
Db        934 VQMAFVKLAALTGTYVYDHLAPLQHWAHTGLRDLAVAVEPVVFSDMETKIITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||||||: ||||||||||||||||||||||||||||||||||||||
Db        994 GDIILGLPVSARRGREILLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db       1054 QVEGEVQVVSTATQSFLATCINGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1114 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDGRGSLLSPRPVSYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||||||||||||||
Db       1174 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206


RESULT 7
Q9J3H3
ID   Q9J3H3      PRELIMINARY;      PRT;  3010 AA.
AC   Q9J3H3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD19;
RA   Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT   "Characteristics of hepatitis C viral genome associated with disease
RT   progression.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF207760; AAF65950.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327234 MW;  44C34677649CB8DD CRC64;

  Query Match             96.0%;  Score 1971;  DB 12;  Length 3010;
  Best Local Similarity   93.9%;  Pred. No. 9.7e-156;
  Matches  369;  Conservative   14;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||||||||||||:||| |||||||||| |||||||| ||:||||||||||||
Db        814 MAASCGGAVFIGLALLTLSPHYKVFLARLIWWLQYFITRVEAHLHVWVPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              |||||||||||| |||:|||||||||||||||||:|||||||||||||:| || |||||
Db        874 ILLTCAVHPELIFSITKILLAIFGPLMVLQAGITRVPYFVRAQGLIRACLLARKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||||||||||||||||||||||||||||||||||||||:|||| ||||||||||||
Db        934 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||:|| |||||| ||||||||||||||||||||||||||||||:|
Db        994 GDIILGLPVSARRGKEIFLGPADNLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||:||||:|||||||||||||||||||||||||||||:
```

```
Db       1054 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGAKTLAGPKGPITQMYTNVDQDLVGWQS 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db       1114 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
Db       1174 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 1206


RESULT 8
Q68788
ID   Q68788      PRELIMINARY;      PRT;  3010 AA.
AC   Q68788;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   HCV polyprotein (Genome polyprotein).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96362158; PubMed=8720135;
RA   Seki M., Honda Y.;
RT   "Phosphorothioate antisense oligodeoxynucleotides capable of
RT   inhibiting Hepatitis C virus gene expression: In vitro translation
RT   assay.";
RL   J. Biochem. 118:1199-1204(1995).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC      LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC      PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC      PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; D45172; BAA08120.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  326880 MW;  EED840E6A050E766 CRC64;

  Query Match            96.0%;  Score 1970;  DB 12;  Length 3010;
  Best Local Similarity  94.1%;  Pred. No. 1.2e-155;
  Matches  370;  Conservative   13;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| |||||||||| ||:|||||||||| |||||||||||||||||||
Db        814 MAASCGGAVFVGLVLLTLSPYYKVFLAKLIWWLQYLITRAEAHLQVWIPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              |||||||||||||||||||| ||||||||||:|||||||||||||||| ||||||
Db        874 ILLTCAVHPELIFDITKLLLAILGPLMVLQAGITRVPYFVRAQGLIRACMLVWKAAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||:|||||||||:|||||||:|||||||||||||||||:||||| ||||||||||
Db        934 VQMAFVKLAALTGTYIYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||:|||||||: ||||||||||||||||||||||||||||||||||
Db        994 GDIILGLPVSARRGKEILLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              |||||||||||||||||||:|||||||:|||||||||||||||||||||||||||||| |
Db       1054 QVEGEVQVVSTATQSFLATCINGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWPA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db       1114 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              || |||||||||||||||||:|||||||||||
Db       1174 GHVVGIFRAAVCTRGVAKAVDFVPVESMETTMR 1206


RESULT 9
Q81755
ID   Q81755      PRELIMINARY;      PRT;  1186 AA.
AC   Q81755;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Polyprotein (Fragment).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91088550; PubMed=2175903;
RA   Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA   Sugimura T., Shimotohno K.;
RT   "Molecular cloning of the human hepatitis C virus genome from Japanese
RT   patients with non-A, non-B hepatitis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92044440; PubMed=1658196;
RA   Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,
RA   Miyakawa Y., Mayumi M.;
RT   "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT   from a human carrier: comparison with reported isolates for conserved
```

```
RT   and divergent regions.";
RL   J. Gen. Virol. 72:2697-2704(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91140698; PubMed=1847440;
RA   Takamizawa A., Mori C., Manabe S., Murakami S., Fujita J., Onishi E.,
RA   Andoh T., Yoshida I., Okayama H.;
RT   "The structure and organization of the Hepatitis C virus genome
RT   isolated from human carriers.";
RL   J. Virol. 65:1105-1113(1991).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91172826; PubMed=1848704;
RA   Choo Q.-L., Richman K., Han J.H., Berger K., Lee C., Dong C.,
RA   Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.,
RA   Bradley D.W., Kuo G., Houghton M.;
RT   "Genetic organization and diversity of the hepatitis C virus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92230206; PubMed=1314449;
RA   Chen P., Lin M., Tai K., Liu P., Lin C., Chen D.;
RT   "The Taiwanese hepatitis C virus genome: Sequence determination and
RT   mapping the 5'termini of viral genomic and antigenomic RNA.";
RL   Virology 188:102-113(1992).
RN   [6]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92230232; PubMed=1314459;
RA   Okamoto H., Kurai K., Okada S., Yamamoto K., Iizuka H., Tanaka T.,
RA   Fukuda S., Tsuda F., Mishiro S.;
RT   "Full-length sequence of a hepatitis C virus genome having poor
RT   homology to reported isolates: Comparative study of four distinct
RT   genotypes.";
RL   Virology 188:331-341(1992).
RN   [7]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93323208; PubMed=8392606;
RA   Hijikata M., Mizushima H., Akagi T., Mori S., Kakiuchi N., Kato N.,
RA   Tanaka T., Kimura K., Shimotohno K.;
RT   "Two distinct proteinase activities required for the processing of a
RT   putative nonstructural precursor protein of hepatitis C virus.";
RL   J. Virol. 67:4665-4675(1993).
RN   [8]
RP   SEQUENCE FROM N.A.
RA   Hijikata M.;
RL   Submitted (DEC-1993) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; D11397; BAA20975.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PS0329; PS0329.
DR   PDB; 1DXP; 28-MAR-02.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   ATP-binding; Helicase; Hydrolase.
FT   NON_TER       1       1
SQ   SEQUENCE   1186 AA;  126280 MW;  34170478BA23729A CRC64;
```

```
  Query Match           95.9%;  Score 1968;  DB 12;  Length 1186;
  Best Local Similarity   94.4%;  Pred. No. 5.2e-156;
  Matches  371;  Conservative   11;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| |||||||||| |||||||||| ||| |||||||:|||||||||||
Db         92 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFITRAEAHLQVWVPPLNVRGGRDAI 151

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              |||||||||||||||||||| |||||||||:|:|||||||||||||||||||| |||||
Db        152 ILLTCAVHPELIFDITKLLLAILGPLMVLQAGMTRVPYFVRAQGLIRACMLVRKVAGGHY 211

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              ||||||||||||||||||||||:|||||||||||||||:||||| ||||||||||||
Db        212 VQMAFMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 271

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||||||||||||:|||||||:|  ||||||||||||||||||||||||||||||||||
Db        272 GDIISGLPVSARRGKEILLGPADSFGEQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 331

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||:|||||:||||||||||||||||:|||||||||||||||||||||||||||||||| |
Db        332 QVDGEVQVLSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWPA 391

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        392 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 451

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              || |||||||||||||||||||||||||||||||
Db        452 GHVVGIFRAAVCTRGVAKAVDFIPVESMETTMR 484
```

```
RESULT 10
Q81817
ID   Q81817      PRELIMINARY;      PRT;  2284 AA.
AC   Q81817;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Polyprotein precursor (Genome polyprotein).
OS   Hepatitis C virus type 2.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=40271;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94068484; PubMed=7504283;
RA   Hijikata M., Mizushima H., Tanji Y., Komada Y., Hirowatari Y.,
RA   Akagi T., Kimura K., Shimotohno K.;
RT   "Proteolytic processing and membrane association of putative
RT   nonstructural proteins of hepatitia C virus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:10773-10777(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94333810; PubMed=8056334;
RA   Tanji Y., Hijikata M., Hirowatari Y., Shimotohno K.;
RT   "Identification of the domain required for trans-cleavage activity of
RT   hepatitis C viral serine proteinase.";
RL   Gene 145:215-219(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95056078; PubMed=7966638;
RA   Tanji Y., Hizikata M., Hirowatari Y., Shimotohono K.;
RT   "Hepatitis C virus polyprotein processing: kinetics and mutagenic
RT   analysis of serine proteinase-dependent cleavage.";
RL   J. Virol. 68:8418-8422(1994).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95156583; PubMed=7853491;
RA   Tanji Y., Hijikata M., Satoh S., Kaneko T., Shimotohno K.;
RT   "Hepatitis C virus-encoded nonstructual protein NS4A has versatile
```

```
RT   functions in viral protein processing.";
RL   J. Virol. 69:1575-1581(1995).
DR   EMBL; D16435; BAA03905.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;
KW   Signal; Transferase.
FT   SIGNAL       4     20
FT   SIGNAL      55     83
FT   CHAIN       21     54
FT   CHAIN       84    300       NS2.
FT   CHAIN      301    931       NS3.
FT   CHAIN      932    985       NS4A.
FT   CHAIN      986   1246       NS4B.
FT   CHAIN     1247   1693       NS5A.
FT   CHAIN     1694   2284       NS5B.
SQ   SEQUENCE   2284 AA;  247213 MW;  DC272A1517046337 CRC64;

  Query Match           95.9%;  Score 1968;  DB 12;  Length 2284;
  Best Local Similarity   94.4%;  Pred. No. 1.2e-155;
  Matches  371;  Conservative   11;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||:|| |||||||||| |||||||||| ||| |||||||:||||||||||||
Db         88 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFITRAEAHLQVWVPPLNVRGGRDAI 147

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              |||||||||||||||||||| ||||||||:|:|||||||||||||||||| |||||
Db        148 ILLTCAVHPELIFDITKLLLAILGPLMVLQAGMTRVPYFVRAQGLIRACMLVRKVAGGHY 207

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||||||||||||||:|||||||||||||||:||||| ||||||||||||
Db        208 VQMAFMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 267

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||||||||||||:|||||||:|  ||||||||||||||||||||||||||||||||||
Db        268 GDIISGLPVSARRGKEILLGPADSFGEQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 327

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||:|||||:|||||||||||||||:|||||||||||||||||||||||||||||||| |
Db        328 QVDGEVQVLSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWPA 387

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        388 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 447

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              || |||||||||||||||||||||||||||||
Db        448 GHVVGIFRAAVCTRGVAKAVDFIPVESMETTMR 480


RESULT 11
P89966
ID   P89966      PRELIMINARY;      PRT;  3010 AA.
AC   P89966;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   RNA for polyprotein (Genome polyprotein).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=type 1b;
RA   Tanaka T.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=type 1b;
RA   TANAKA T.;
RT   "TMORF.";
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; D89872; BAA14035.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PQ0804; PQ0804.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
```

```
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327023 MW;  E075BD9CFD8D1261 CRC64;

  Query Match            95.9%;  Score 1968;  DB 12;  Length 3010;
  Best Local Similarity  94.4%;  Pred. No. 1.7e-155;
  Matches  371;  Conservative   11;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| |||||||||| |||||||||| ||| |||||||:||||||||||||
Db        814 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFITRAEAHLQVWVPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              |||||||||||||||||||| |||||||||:|:|||||||||||||||||| |||||
Db        874 ILLTCAVHPELIFDITKLLLAILGPLMVLQAGMTRVPYFVRAQGLIRACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              ||||||||||||||||||||:|||||||||||||||||:||||| ||||||||||||
Db        934 VQMAFMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||||||||||||:||||||||:|  ||||||||||||||||||||||||||||||||||
Db        994 GDIISGLPVSARRGKEILLGPADSFGEQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||:|||||:|||||||||||||||||:|||||||||||||||||||||||||||||||| |
Db       1054 QVDGEVQVLSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWPA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1114 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              || |||||||||||||||||||||||||||||
Db       1174 GHVVGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206


RESULT 12
Q9J3G6
ID   Q9J3G6      PRELIMINARY;      PRT;  3010 AA.
AC   Q9J3G6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD26;
RA   Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT   "Characteristics of hepatitis C viral genome associated with disease
RT   progression.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF207767; AAF65957.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PQ0254; PQ0254.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   MEROPS; S29.002; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327165 MW;  74FAB6B80F24837B CRC64;

  Query Match            95.9%;  Score 1968;  DB 12;  Length 3010;
  Best Local Similarity  94.7%;  Pred. No. 1.7e-155;
  Matches  372;  Conservative    7;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||| ||:||||||||||||| |||||||||| ||| |||||||||||||||||||
Db        814 MAASCGGGVFVGLALLTLSPYYKVFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| |||||||||||||||||||||||||| ||:|||||||||||||||||| |||||
Db        874 ILLVCAVHPELIFDITKLLLAIFGPLMVLQASITRVPYFVRAQGLIRACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| :|||||||||||||| ||| |||||||||||||||:||||| ||||||||||||
Db        934 VQMALVKLAALTGTYVYDHLAPLQHWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
```

```
                ||||| |||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db        994 GDIILGLPVSARRGREILLGPADSFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1054 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
                |||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db       1114 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPL 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
                || |||||||||||||||||:||||||||||
Db       1174 GHVVGIFRAAVCTRGVAKAVDFVPVESMETTMR 1206


RESULT 13
Q99AU2
ID   Q99AU2      PRELIMINARY;      PRT;  3010 AA.
AC   Q99AU2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus type 1b.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=31647;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=chimera of HCV-BK;
RA   Thomson M., Nascimbeni M., Gonzales S., Murthy K., Rehermann B.,
RA   Liang J.;
RT   "Analyses of viral sequences and virus-specific immune responses
RT   during serial passage of an infectious hepatitis C virus serotype 1b
RT   clone in chimpanzees.";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF333324; AAK08509.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PQ0804; PQ0804.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1NS3.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327007 MW;  053B9A653B0AB335 CRC64;

  Query Match             95.8%;  Score 1967;  DB 12;  Length 3010;
  Best Local Similarity   93.9%;  Pred. No. 2.1e-155;
  Matches  369;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||||:|| |||||||||| |||||||||| ||| |||||||:||||||||||||
Db        814 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFITRAEAHLQVWVPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              |||||||||||||||||||||| ||||||||||:|||||||||| ||||||| |||||
Db        874 ILLTCAVHPELIFDITKLLLAILGPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||| |||||:|:|||||:|||||||||||||||||:||||| ||||||||||||
Db        934 VQMAFMKLGALTGTYIYNHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||:||||||||: ||:|||||||||||||||||||||||||||||||||
Db        994 GDIILGLPVSARRGKEILLGPADSLEGRGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||
Db       1054 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1114 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||||:||||||||||
Db       1174 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 1206


RESULT 14
Q9DTD7
ID   Q9DTD7      PRELIMINARY;      PRT;  3010 AA.
AC   Q9DTD7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCVT217;
RA   Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA   Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA   Mishiro S.;
```

```
RT   "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT   with hepatocellular carcinoma: the 'progression score' revisited.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AB049100; BAB18813.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0804; PQ0804.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   MEROPS; S29.002; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  326909 MW;  5505C62EB0DA0519 CRC64;

  Query Match             95.8%;  Score 1967;  DB 12;  Length 3010;
  Best Local Similarity   94.4%;  Pred. No. 2.1e-155;
  Matches  371;  Conservative   11;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              :|||||||||||||||||||||| |||||||||| ||| ||:||||:||||||||||||
Db        814 VAASCGGAVFIGLALLTLSPYYKVFLARLIWWLQYFITRAEAYLQVWVPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              |||||||||||||||||||||||||:|||||||:||||| ||||||||||||||||||||

Db        874 ILLTCAVHPELIFDITKLLLAIFGPIMVLQAGITRVPYFVPAQGLIRACMLVRKAAGGHY 933

Qy        121 VQMAFMKLAAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              :||| :|||||||||||||| ||| |||||||||||||||:|||| |:||||||||||
Db        934 IQMALVKLAALTGTYVYDHLAPLQHWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||||||: ||||||||||||||||||||||||||||||||||||||
Db        994 GDIILGLPVSARRGREILLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db       1054 QVEGEVQVVSTATQSFLATCINGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       1114 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDGRGSLLSPRPVSYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||||||||||||||
Db       1174 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206


RESULT 15
Q9QIX6
ID   Q9QIX6      PRELIMINARY;      PRT;  3010 AA.
AC   Q9QIX6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Genome polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD8-1;
RX   MEDLINE=20013325; PubMed=10544098;
RA   Nagayama K., Kurosaki M., Enomoto N., Maekawa S.y., Miyasaka Y.,
RA   Tazawa J.i., Izumi N., Marumo F., Sato C.;
RT   "Time-related changes in full-length hepatitis C virus and hepatitis
RT   activity.";
RL   Virology 263:244-253(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD8-1;
RA   Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,
RA   Sakamoto N., Fukuma T., Tazawa J., Izumi N., Marumo F., Sato C.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA (BY SIMILARITY).
DR   EMBL; AF165059; AAD56194.1; -.
DR   PIR; A61196; A61196.
DR   PIR; PQ0246; PQ0246.
DR   PIR; PQ0254; PQ0254.
DR   PIR; PQ0804; PQ0804.
DR   PIR; PS0329; PS0329.
DR   HSSP; P26663; 1JXP.
DR   MEROPS; S29.002; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
```

```
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ   SEQUENCE   3010 AA;  327466 MW;  4613744EC4D4A013 CRC64;

  Query Match            95.8%;  Score 1967;  DB 12;  Length 3010;
  Best Local Similarity  94.4%;  Pred. No. 2.1e-155;
  Matches  371;  Conservative    8;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:||| |||||||||| |||||||||| ||| |||||||:||||||||||||
Db        814 MAASCGGAVFVGLAFLTLSPYYKVFLARLIWWLQYFITRAEAHLQVWVPPLNVRGGRDAI 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| | ||||||||||||||||||||||||||:||||||| |||||||||| |||||
Db        874 ILLMCVVHPELIFDITKLLLAIFGPLMVLQAGITRVPYFVRAHGLIRACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||:||||||||||||| ||| |||:||||||||||:||||| ||||||||||||
Db        934 VQMAFVKLAALTGTYVYDHLAPLQHWAHSGLRDLAVAVEPVVFSDMETKIITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||||||: |||||||||||||||||||||||||||||||||||||
Db        994 GDIILGLPVSARRGREILLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db       1054 QVEGEVQVVSTATQSFLATCINGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       1114 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDGRGSLLSPRPVSYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||||||||||||||
Db       1174 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206


  Search completed: May  6, 2004, 09:35:46
  Job time : 37.7566 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:25:16 ; Search time 15.4246 Seconds
(without alignments)
1315.364 Million cell updates/sec

Title: US-10-650-585-11
Perfect score: 2053
Sequence: 1 MAASCGGAVFIGLALLTLSP.....RGVAKAVDFIPVESMETTMR 393

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1951	95.0	2201	4	US-09-539-601-6	Sequence 6, Appli
2	1951	95.0	2201	4	US-09-539-601-15	Sequence 15, Appl
3	1951	95.0	3010	4	US-09-539-601-3	Sequence 3, Appli
4	1951	95.0	3010	4	US-09-539-601-21	Sequence 21, Appl
5	1951	95.0	3010	4	US-09-539-601-27	Sequence 27, Appl
6	1946	94.8	1692	3	US-09-263-933-4	Sequence 4, Appli
7	1946	94.8	1692	4	US-09-919-901-4	Sequence 4, Appli
8	1946	94.8	2307	3	US-09-263-933-2	Sequence 2, Appli
9	1946	94.8	2307	4	US-09-919-901-2	Sequence 2, Appli
10	1944	94.7	3010	4	US-09-539-601-33	Sequence 33, Appl
11	1943	94.6	1692	3	US-09-263-933-11	Sequence 11, Appl
12	1943	94.6	1692	4	US-09-919-901-11	Sequence 11, Appl
13	1943	94.6	2307	3	US-09-263-933-9	Sequence 9, Appli
14	1943	94.6	2307	4	US-09-919-901-9	Sequence 9, Appli
15	1934	94.2	1692	3	US-09-263-933-18	Sequence 18, Appl
16	1934	94.2	1692	4	US-09-919-901-18	Sequence 18, Appl
17	1934	94.2	2307	3	US-09-263-933-16	Sequence 16, Appl
18	1934	94.2	2307	4	US-09-919-901-16	Sequence 16, Appl
19	1928	93.9	3010	3	US-09-014-416-3	Sequence 3, Appli
20	1888	92.0	2013	1	US-08-324-977-12	Sequence 12, Appl
21	1888	92.0	2013	2	US-08-384-616-12	Sequence 12, Appl
22	1888	92.0	2013	2	US-08-904-686A-12	Sequence 12, Appl
23	1888	92.0	2013	3	US-09-315-850-12	Sequence 12, Appl
24	1888	92.0	2201	4	US-08-952-981A-2	Sequence 2, Appli
25	1888	92.0	2620	1	US-08-324-977-32	Sequence 32, Appl
26	1888	92.0	2620	2	US-08-384-616-32	Sequence 32, Appl
27	1888	92.0	2620	2	US-08-904-686A-32	Sequence 32, Appl
28	1888	92.0	2620	3	US-09-315-850-32	Sequence 32, Appl
29	1888	92.0	2621	1	US-08-324-977-36	Sequence 36, Appl
30	1888	92.0	2621	2	US-08-384-616-36	Sequence 36, Appl
31	1888	92.0	2621	2	US-08-904-686A-36	Sequence 36, Appl
32	1888	92.0	2621	3	US-09-315-850-36	Sequence 36, Appl
33	1888	92.0	3010	1	US-08-324-977-2	Sequence 2, Appli
34	1888	92.0	3010	1	US-08-324-977-14	Sequence 14, Appl
35	1888	92.0	3010	2	US-08-384-616-2	Sequence 2, Appli
36	1888	92.0	3010	2	US-08-384-616-14	Sequence 14, Appl
37	1888	92.0	3010	2	US-08-904-686A-2	Sequence 2, Appli
38	1888	92.0	3010	2	US-08-904-686A-14	Sequence 14, Appl
39	1888	92.0	3010	3	US-09-315-850-2	Sequence 2, Appli
40	1888	92.0	3010	3	US-09-315-850-14	Sequence 14, Appl
41	1767	86.1	1648	1	US-08-188-281B-12	Sequence 12, Appl
42	1767	86.1	1648	5	PCT-US94-07280-12	Sequence 12, Appl
43	1767	86.1	1648	5	PCT-US95-01087-12	Sequence 12, Appl
44	1767	86.1	3011	1	US-08-188-281B-1	Sequence 1, Appli
45	1767	86.1	3011	1	US-08-453-552-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-539-601-6
; Sequence 6, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartenschlager, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2201
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-539-601-6

Query Match 95.0%; Score 1951; DB 4; Length 2201;
Best Local Similarity 93.1%; Pred. No. 8.8e-187;
Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

```
Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||||:|| |||||:|: |||||||||| ||| |||||||||||||||||:
Db          5 MAASCGGAVFVGLILLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAV 64

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||:|||||| |||:|||| |||||||||||||||||| |||||||||| |||||
Db         65 ILLTCAIHPELIFTITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHY 124

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| |||||||||||||||:|||||||||||||||:||||| |:||||||||||
Db        125 VQMALMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAAC 184

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||| |||||: |||||||||||||||||||||||||||||||:|
Db        185 GDIILGLPVSARRGREIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRN 244

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              |||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
Db        245 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 304

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 364

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
```

```
Db        365 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 397


RESULT 2
US-09-539-601-15
; Sequence 15, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
;  APPLICANT: Bartenschlager, Ralf FW
;  TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
;  FILE REFERENCE: all sequences
;  CURRENT APPLICATION NUMBER: US/09/539,601C
;  CURRENT FILING DATE: 2001-08-30
;  EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
;  EARLIER FILING DATE: 1999-04-03
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
;   LENGTH: 2201
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
US-09-539-601-15

  Query Match           95.0%;  Score 1951;  DB 4;  Length 2201;
  Best Local Similarity  93.1%;  Pred. No. 8.8e-187;
  Matches  366;  Conservative   14;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||:|| ||||||:||: |||||||||| ||| ||||||||||||||||||:
Db          5 MAASCGGAVFVGLILLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAV 64

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||:|||||| |||:|||| |||||||||||||||||| |||||||||| |||||
Db         65 ILLTCAIHPELIFTITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHY 124

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| |||||||||||||||:||||||||||||||:||||| |:||||||||||
Db        125 VQMALMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAAC 184

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||| |||||: |||||||||||||||||||||||||||||||:|
Db        185 GDIILGLPVSARRGREIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRN 244

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db        245 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 304

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 364

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
Db        365 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 397


RESULT 3
US-09-539-601-3
; Sequence 3, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
;  APPLICANT: Bartenschlager, Ralf FW
;  TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
;  FILE REFERENCE: all sequences
;  CURRENT APPLICATION NUMBER: US/09/539,601C
;  CURRENT FILING DATE: 2001-08-30
;  EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
;  EARLIER FILING DATE: 1999-04-03
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 3010
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
US-09-539-601-3

  Query Match           95.0%;  Score 1951;  DB 4;  Length 3010;
  Best Local Similarity  93.1%;  Pred. No. 1.4e-186;
  Matches  366;  Conservative   14;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||:|| ||||||:||: |||||||||| ||| ||||||||||||||||||:
Db        814 MAASCGGAVFVGLILLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAV 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||:|||||| |||:|||| |||||||||||||||||| |||||||||| |||||
Db        874 ILLTCAIHPELIFTITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| |||||||||||||||:||||||||||||||:||||| |:||||||||||
Db        934 VQMALMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||| |||||: |||||||||||||||||||||||||||||||:|
Db        994 GDIILGLPVSARRGREIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db       1054 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1114 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
Db       1174 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 1206


RESULT 4
US-09-539-601-21
; Sequence 21, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
;  APPLICANT: Bartenschlager, Ralf FW
;  TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
;  FILE REFERENCE: all sequences
;  CURRENT APPLICATION NUMBER: US/09/539,601C
;  CURRENT FILING DATE: 2001-08-30
;  EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
;  EARLIER FILING DATE: 1999-04-03
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
;   LENGTH: 3010
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
US-09-539-601-21

  Query Match           95.0%;  Score 1951;  DB 4;  Length 3010;
  Best Local Similarity  93.1%;  Pred. No. 1.4e-186;
  Matches  366;  Conservative   14;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||:|| ||||||:||: |||||||||| ||| ||||||||||||||||||:
Db        814 MAASCGGAVFVGLILLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAV 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||:|||||| |||:|||| |||||||||||||||||| |||||||||| |||||
Db        874 ILLTCAIHPELIFTITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| |||||||||||||||:||||||||||||||:||||| |:||||||||||
Db        934 VQMALMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
```

```
Db        994 GDIILGLPVSARRGREIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300

Db       1054 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360

Db       1114 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393

Db       1174 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 1206


RESULT 5
US-09-539-601-27
; Sequence 27, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
;  APPLICANT: Bartenschlager, Ralf FW
;  TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
;  FILE REFERENCE: all sequences
;  CURRENT APPLICATION NUMBER: US/09/539,601C
;  CURRENT FILING DATE: 2001-08-30
;  EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
;  EARLIER FILING DATE: 1999-04-03
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
;   LENGTH: 3010
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
US-09-539-601-27

  Query Match             95.0%;  Score 1951;  DB 4;  Length 3010;
  Best Local Similarity   93.1%;  Pred. No. 1.4e-186;
  Matches  366;  Conservative   14;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60

Db        814 MAASCGGAVFVGLILLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAV 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120

Db        874 ILLTCAIHPELIFTITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180

Db        934 VQMALMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240

Db        994 GDIILGLPVSARRGREIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300

Db       1054 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360

Db       1114 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393

Db       1174 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 1206


RESULT 6
US-09-263-933-4
; Sequence 4, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT: Jackson, Roberta L.
;  APPLICANT: Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/263,933
;  CURRENT FILING DATE: 1999-03-08
;  EARLIER APPLICATION NUMBER: 09/129,611
;  EARLIER FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 1692
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
US-09-263-933-4

  Query Match             94.8%;  Score 1946;  DB 3;  Length 1692;
  Best Local Similarity   92.9%;  Pred. No. 1.9e-186;
  Matches  365;  Conservative   13;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60

Db         93 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFTTRAEAHLHVWIPPLNARGGRDAI 152

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120

Db        153 ILLMCAVHPELIFDITKLLIAILGPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHY 212

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180

Db        213 VQMAFMKLGALTGTYIYNHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 272

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240

Db        273 GDIILGLPVSARRGKEILLGPADSLEGRGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 332

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300

Db        333 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 392

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360

Db        393 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 452

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393

Db        453 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 485


RESULT 7
US-09-919-901-4
; Sequence 4, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT:  Jackson, Roberta L.
;  APPLICANT:  Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION:  OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/919,901
;  CURRENT FILING DATE:  2001-08-02
;  PRIOR APPLICATION NUMBER: 09/263,933
;  PRIOR FILING DATE: 1999-02-08
;  PRIOR APPLICATION NUMBER: 09/129,611
;  PRIOR FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 1692
```

```
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: :
US-09-919-901-4

  Query Match             94.8%;  Score 1946;  DB 4;  Length 1692;
  Best Local Similarity   92.9%;  Pred. No. 1.9e-186;
  Matches  365;  Conservative   13;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| |||||||||| |||||||||   || |||| ||||||| |||||||
Db         93 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFTTRAEAHLHVWIPPLNARGGRDAI 152

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| |||||||||||||:|| |||||||||:|||||||||| ||||||| |||||
Db        153 ILLMCAVHPELIFDITKLLIAILGPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHY 212

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||| ||||||:|:|||||:||||||||||||||||:||||| ||||||||||||
Db        213 VQMAFMKLGALTGTYIYNHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 272

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||:|||||||: ||:||||||||||||||||||||||||||||||||
Db        273 GDIILGLPVSARRGKEILLGPADSLEGRGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 332

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db        333 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 392

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        393 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 452

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||||:||||||||||
Db        453 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 485


RESULT 8
US-09-263-933-2
; Sequence 2, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT: Jackson, Roberta L.
;  APPLICANT: Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/263,933
;  CURRENT FILING DATE: 1999-03-08
;  EARLIER APPLICATION NUMBER: 09/129,611
;  EARLIER FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 2307
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
US-09-263-933-2

  Query Match             94.8%;  Score 1946;  DB 3;  Length 2307;
  Best Local Similarity   92.9%;  Pred. No. 3e-186;
  Matches  365;  Conservative   13;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| |||||||||| |||||||||   || |||| ||||||| |||||||
Db       185 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFTTRAEAHLHVWIPPLNARGGRDAI 244

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| |||||||||||||:|| |||||||||:|||||||||| ||||||| |||||
Db       245 ILLMCAVHPELIFDITKLLIAILGPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHY 304

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||| ||||||:|:|||||:||||||||||||||||:||||| ||||||||||||
Db       305 VQMAFMKLGALTGTYIYNHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 364

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||:|||||||: ||:||||||||||||||||||||||||||||||||
Db       365 GDIILGLPVSARRGKEILLGPADSLEGRGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 424

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db       425 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 484

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       485 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 544

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||||:||||||||||
Db       545 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 577


RESULT 9
US-09-919-901-2
; Sequence 2, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT:  Jackson, Roberta L.
;  APPLICANT:  Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION:  OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/919,901
;  CURRENT FILING DATE:  2001-08-02
;  PRIOR APPLICATION NUMBER: 09/263,933
;  PRIOR FILING DATE: 1999-02-08
;  PRIOR APPLICATION NUMBER: 09/129,611
;  PRIOR FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 2307
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: :
US-09-919-901-2

  Query Match             94.8%;  Score 1946;  DB 4;  Length 2307;
  Best Local Similarity   92.9%;  Pred. No. 3e-186;
  Matches  365;  Conservative   13;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| |||||||||| |||||||||   || |||| ||||||| |||||||
Db       185 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFTTRAEAHLHVWIPPLNARGGRDAI 244

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| |||||||||||||:|| |||||||||:|||||||||| ||||||| |||||
Db       245 ILLMCAVHPELIFDITKLLIAILGPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHY 304

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||| ||||||:|:|||||:||||||||||||||||:||||| ||||||||||||
Db       305 VQMAFMKLGALTGTYIYNHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 364

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||:|||||||: ||:||||||||||||||||||||||||||||||||
Db       365 GDIILGLPVSARRGKEILLGPADSLEGRGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 424

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db       425 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 484
```

```
Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 544

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
Db        545 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 577


RESULT 10
US-09-539-601-33
; Sequence 33, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
;  APPLICANT: Bartenschlager, Ralf FW
;  TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
;  FILE REFERENCE: all sequences
;  CURRENT APPLICATION NUMBER: US/09/539,601C
;  CURRENT FILING DATE: 2001-08-30
;  EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
;  EARLIER FILING DATE: 1999-04-03
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
;   LENGTH: 3010
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
US-09-539-601-33

  Query Match             94.7%;  Score 1944;  DB 4;  Length 3010;
  Best Local Similarity   92.9%;  Pred. No. 7.1e-186;
  Matches  365;  Conservative   14;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| ||||||:||: |||||||||| ||| ||||||||||||||||||:
Db        814 MAASCGGAVFVGLILLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAV 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||:|||||| |||:|||| |||||||||||||||| |||||||||| |||||
Db        874 ILLTCAIHPELIFTITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| |||||||||||||||:|||||||||||||||:||||| |:||||||||||
Db        934 VQMALMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||| |||||: |||||||||||||||||||||||||||||:|
Db        994 GDIILGLPVSARRGREIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
Db       1054 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1114 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||| ||||
Db       1174 GHAVGIFRAAVCTRGVAKAVDFVPVESMGTTMR 1206


RESULT 11
US-09-263-933-11
; Sequence 11, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT: Jackson, Roberta L.
;  APPLICANT: Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/263,933
;  CURRENT FILING DATE: 1999-03-08
;  EARLIER APPLICATION NUMBER: 09/129,611
;  EARLIER FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 1692
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
US-09-263-933-11

  Query Match             94.6%;  Score 1943;  DB 3;  Length 1692;
  Best Local Similarity   92.6%;  Pred. No. 3.8e-186;
  Matches  364;  Conservative   14;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| |||||||||| |||||||||  || |||| ||||||| |||||||
Db         93 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFTTRAEAHLHVWIPPLNARGGRDAI 152

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| |||||||||||||:|| ||||||||||:|||||||||| |||||| |||||
Db        153 ILLMCAVHPELIFDITKLLIAILGPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHY 212

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||| |||||:|:||||||:||||||||||||||||:|||| ||||||||||||
Db        213 VQMAFMKLGALTGTYIYNHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 272

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||:|||||||: ||:|||||||||||||||||||||||||||||||||
Db        273 GDIILGLPVSARRGKEILLGPADSLEGRGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 332

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
Db        333 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 392

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        393 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSAGGPLLCPS 452

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
Db        453 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 485


RESULT 12
US-09-919-901-11
; Sequence 11, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT:  Jackson, Roberta L.
;  APPLICANT:  Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION:  OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/919,901
;  CURRENT FILING DATE:  2001-08-02
;  PRIOR APPLICATION NUMBER: 09/263,933
;  PRIOR FILING DATE: 1999-02-08
;  PRIOR APPLICATION NUMBER: 09/129,611
;  PRIOR FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 1692
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: :
US-09-919-901-11

  Query Match             94.6%;  Score 1943;  DB 4;  Length 1692;
```

```
  Best Local Similarity   92.6%;  Pred. No. 3.8e-186;
  Matches  364;  Conservative   14;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| |||||||||| |||||||||  || |||| ||||||| |||||||
Db         93 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFTTRAEAHLHVWIPPLNARGGRDAI 152

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| |||||||||||||:|| |||||||||:|||||||||| ||||||| |||||
Db        153 ILLMCAVHPELIFDITKLLIAILGPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHY 212

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||| |||||:|:|||||:||||||||||||||||:||||| ||||||||||||
Db        213 VQMAFMKLGALTGTYIYNHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 272

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||:||||||||: ||:||||||||||||||||||||||||||||||||
Db        273 GDIILGLPVSARRGKEILLGPADSLEGRGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 332

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db        333 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 392

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:|||||||||||||||||||||||||||||||||||||||||:||||||||
Db        393 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSAGGPLLCPS 452

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
Db        453 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 485


RESULT 13
US-09-263-933-9
; Sequence 9, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT: Jackson, Roberta L.
;  APPLICANT: Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/263,933
;  CURRENT FILING DATE: 1999-03-08
;  EARLIER APPLICATION NUMBER: 09/129,611
;  EARLIER FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 2307
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
US-09-263-933-9

  Query Match             94.6%;  Score 1943;  DB 3;  Length 2307;
  Best Local Similarity   92.6%;  Pred. No. 6e-186;
  Matches  364;  Conservative   14;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| |||||||||| |||||||||  || |||| ||||||| |||||||
Db        185 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFTTRAEAHLHVWIPPLNARGGRDAI 244

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| |||||||||||||:|| |||||||||:|||||||||| ||||||| |||||
Db        245 ILLMCAVHPELIFDITKLLIAILGPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHY 304

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||| |||||:|:|||||:||||||||||||||||:||||| ||||||||||||
Db        305 VQMAFMKLGALTGTYIYNHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 364

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||:||||||||: ||:||||||||||||||||||||||||||||||||
Db        365 GDIILGLPVSARRGKEILLGPADSLEGRGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 424

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db        425 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 484

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:|||||||||||||||||||||||||||||||||||||||||:||||||||
Db        485 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSAGGPLLCPS 544

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
Db        545 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 577


RESULT 14
US-09-919-901-9
; Sequence 9, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT:  Jackson, Roberta L.
;  APPLICANT:  Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION:  OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/919,901
;  CURRENT FILING DATE:  2001-08-02
;  PRIOR APPLICATION NUMBER: 09/263,933
;  PRIOR FILING DATE: 1999-02-08
;  PRIOR APPLICATION NUMBER: 09/129,611
;  PRIOR FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 2307
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: :
US-09-919-901-9

  Query Match             94.6%;  Score 1943;  DB 4;  Length 2307;
  Best Local Similarity   92.6%;  Pred. No. 6e-186;
  Matches  364;  Conservative   14;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| |||||||||| |||||||||  || |||| ||||||| |||||||
Db        185 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFTTRAEAHLHVWIPPLNARGGRDAI 244

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| |||||||||||||:|| |||||||||:|||||||||| ||||||| |||||
Db        245 ILLMCAVHPELIFDITKLLIAILGPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHY 304

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||| |||||:|:|||||:||||||||||||||||:||||| ||||||||||||
Db        305 VQMAFMKLGALTGTYIYNHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 364

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||:||||||||: ||:||||||||||||||||||||||||||||||||
Db        365 GDIILGLPVSARRGKEILLGPADSLEGRGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 424

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db        425 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 484

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:|||||||||||||||||||||||||||||||||||||||||:||||||||
Db        485 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSAGGPLLCPS 544

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
Db        545 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 577
```

RESULT 15
US-09-263-933-18
; Sequence 18, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-18

Query Match 94.2%; Score 1934; DB 3; Length 1692;
Best Local Similarity 92.4%; Pred. No. 3.1e-185;
Matches 363; Conservative 14; Mismatches 16; Indels 0; Gaps 0;

```
Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| |||||||||| |||||||||| || |||| ||||||| |||||||
Db         93 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFTTRAEAHLHVWIPPLNARGGRDAI 152

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| |||||||||||||:|| ||||||||||:|||||||||| ||||||| |||||
Db        153 ILLMCAVHPELIFDITKLLIAILGPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHY 212

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||| ||||||:|:|||||:|||||||||||||||:||||| ||||||||||
Db        213 VQMAFMKLGALTGTYIYNHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAA 272

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||:|||||||: ||:||||||||||||||||||||||||||||||||
Db        273 GDIILGLPVSARRGKEILLGPADSLEGRGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 332

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db        333 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 392

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:|||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db        393 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSAGGPLLCPS 452

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||:||||||||||
Db        453 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 485
```

Search completed: May 6, 2004, 09:39:02
Job time : 16.4246 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:30:56 ; Search time 40.2025 Seconds
(without alignments)
2713.357 Million cell updates/sec

Title: US-10-650-585-11
Perfect score: 2053
Sequence: 1 MAASCGGAVFIGLALLTLSP.....RGVAKAVDFIPVESMETTMR 393

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2053	100.0	393	13	US-10-017-736-11	Sequence 11, Appl
2	2053	100.0	393	16	US-10-650-585-11	Sequence 11, Appl
3	2053	100.0	409	13	US-10-017-736-2	Sequence 2, Appli
4	2053	100.0	409	16	US-10-650-585-2	Sequence 2, Appli
5	1987	96.8	380	13	US-10-017-736-12	Sequence 12, Appl
6	1987	96.8	380	16	US-10-650-585-12	Sequence 12, Appl
7	1951	95.0	2201	13	US-10-029-907-3	Sequence 3, Appli
8	1951	95.0	2201	14	US-10-309-561-3	Sequence 3, Appli
9	1951	95.0	3010	12	US-10-467-000-1	Sequence 1, Appli
10	1946	94.8	1692	10	US-09-919-901-4	Sequence 4, Appli
11	1946	94.8	1692	14	US-10-191-966-4	Sequence 4, Appli
12	1946	94.8	2307	10	US-09-919-901-2	Sequence 2, Appli
13	1946	94.8	2307	14	US-10-191-966-2	Sequence 2, Appli
14	1943	94.6	1692	10	US-09-919-901-11	Sequence 11, Appl
15	1943	94.6	1692	14	US-10-191-966-11	Sequence 11, Appl
16	1943	94.6	2307	10	US-09-919-901-9	Sequence 9, Appli
17	1943	94.6	2307	14	US-10-191-966-9	Sequence 9, Appli
18	1934	94.2	1692	10	US-09-919-901-18	Sequence 18, Appl
19	1934	94.2	1692	14	US-10-191-966-18	Sequence 18, Appl
20	1934	94.2	2307	10	US-09-919-901-16	Sequence 16, Appl
21	1934	94.2	2307	14	US-10-191-966-16	Sequence 16, Appl
22	1888	92.0	2201	13	US-10-085-476-2	Sequence 2, Appli
23	1842	89.7	352	13	US-10-017-736-13	Sequence 13, Appl
24	1842	89.7	352	16	US-10-650-585-13	Sequence 13, Appl
25	1778	86.6	341	13	US-10-017-736-14	Sequence 14, Appl
26	1778	86.6	341	16	US-10-650-585-14	Sequence 14, Appl
27	1772	86.3	2985	14	US-10-259-275-40	Sequence 40, Appl
28	1766	86.0	3011	9	US-09-742-659-4	Sequence 4, Appli
29	1766	86.0	3011	10	US-09-891-894-3	Sequence 3, Appli
30	1766	86.0	3011	14	US-10-184-150-3	Sequence 3, Appli
31	1766	86.0	3011	15	US-10-328-997-3	Sequence 3, Appli
32	1766	86.0	3012	9	US-09-238-076-2	Sequence 2, Appli
33	1766	86.0	3012	10	US-09-995-937-2	Sequence 2, Appli
34	1766	86.0	3012	10	US-09-917-563-2	Sequence 2, Appli
35	1764	85.9	3011	9	US-09-916-359-2	Sequence 2, Appli
36	1764	85.9	3011	12	US-10-296-734-406	Sequence 406, App
37	1762	85.8	3011	9	US-09-238-076-20	Sequence 20, Appl
38	1762	85.8	3011	10	US-09-995-937-20	Sequence 20, Appl
39	1762	85.8	3011	10	US-09-917-563-20	Sequence 20, Appl
40	1759	85.7	2894	9	US-09-941-611-23	Sequence 23, Appl
41	1759	85.7	2894	14	US-10-044-995-23	Sequence 23, Appl
42	1759	85.7	3011	9	US-09-952-572-9	Sequence 9, Appli
43	1759	85.7	3011	9	US-09-747-419-20	Sequence 20, Appl
44	1759	85.7	3011	12	US-10-189-359-14	Sequence 14, Appl
45	1759	85.7	3011	14	US-10-259-275-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-10-017-736-11
; Sequence 11, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-11

Query Match 100.0%; Score 2053; DB 13; Length 393;
Best Local Similarity 100.0%; Pred. No. 2.2e-197;
Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120

Qy        121 VQMAFMKLAAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VQMAFMKLAAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||||||||||||||
Db        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393


RESULT 2
US-10-650-585-11
; Sequence 11, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/650,585
;  CURRENT FILING DATE:  2003-08-28
;  PRIOR APPLICATION NUMBER: US/10/017,736A
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 393
;   TYPE: PRT
;   ORGANISM: HCV
US-10-650-585-11

  Query Match             100.0%;  Score 2053;  DB 16;  Length 393;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-197;
  Matches  393;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||||||||||||||
Db        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393


RESULT 3
US-10-017-736-2
; Sequence 2, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/017,736
;  CURRENT FILING DATE:  2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 409
;   TYPE: PRT
;   ORGANISM: HCV
US-10-017-736-2

  Query Match             100.0%;  Score 2053;  DB 13;  Length 409;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-197;
  Matches  393;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 64

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 124

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 184

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 244

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 304

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 364

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||||||||||||||
Db        365 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 397


RESULT 4
US-10-650-585-2
; Sequence 2, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/650,585
;  CURRENT FILING DATE:  2003-08-28
;  PRIOR APPLICATION NUMBER: US/10/017,736A
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 409
;   TYPE: PRT
;   ORGANISM: HCV
US-10-650-585-2

  Query Match             100.0%;  Score 2053;  DB 16;  Length 409;
```

```
  Best Local Similarity   100.0%;  Pred. No. 2.3e-197;
  Matches  393;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 64

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 124

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 184

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 244

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 304

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 364

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||||||||||||||
Db        365 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 397


RESULT 5
US-10-017-736-12
; Sequence 12, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/017,736
;  CURRENT FILING DATE:  2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
;   LENGTH: 380
;   TYPE: PRT
;   ORGANISM: HCV
US-10-017-736-12

  Query Match             96.8%;  Score 1987;  DB 13;  Length 380;
  Best Local Similarity   100.0%;  Pred. No. 9e-191;
  Matches  380;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         14 ALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIF 73
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIF 60

Qy         74 DITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTG 133
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTG 120

Qy        134 TYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 193
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 180

Qy        194 GREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTAT 253
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTAT 240

Qy        254 QSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCG 313
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCG 300

Qy        314 SSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCT 373
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCT 360

Qy        374 RGVAKAVDFIPVESMETTMR 393
              ||||||||||||||||||||
Db        361 RGVAKAVDFIPVESMETTMR 380


RESULT 6
US-10-650-585-12
; Sequence 12, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim (Canada) Ltd.
;  TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
;  FILE REFERENCE: 13/082
;  CURRENT APPLICATION NUMBER: US/10/650,585
;  CURRENT FILING DATE:  2003-08-28
;  PRIOR APPLICATION NUMBER: US/10/017,736A
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: 60/256,031
;  PRIOR FILING DATE: 2000-12-15
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
;   LENGTH: 380
;   TYPE: PRT
;   ORGANISM: HCV
US-10-650-585-12

  Query Match             96.8%;  Score 1987;  DB 16;  Length 380;
  Best Local Similarity   100.0%;  Pred. No. 9e-191;
  Matches  380;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         14 ALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIF 73
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIF 60

Qy         74 DITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTG 133
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTG 120

Qy        134 TYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 193
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 180

Qy        194 GREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTAT 253
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTAT 240

Qy        254 QSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCG 313
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCG 300

Qy        314 SSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCT 373
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCT 360

Qy        374 RGVAKAVDFIPVESMETTMR 393
              ||||||||||||||||||||
Db        361 RGVAKAVDFIPVESMETTMR 380


RESULT 7
US-10-029-907-3
; Sequence 3, Application US/10029907
; Publication No. US20020142350A1
; GENERAL INFORMATION:
;  APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
```

```
;  TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
;  TITLE OF INVENTION:  HEPATITIS C VIRUS
;  FILE REFERENCE: 13/083
;  CURRENT APPLICATION NUMBER: US/10/029,907
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 60/257,857
;  PRIOR FILING DATE: 2000-12-22
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 2201
;   TYPE: PRT
;   ORGANISM: HCV
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: 882
;   OTHER INFORMATION: Xaa is Lys or Arg
;   NAME/KEY: VARIANT
;   LOCATION: 1489
;   OTHER INFORMATION: Xaa is Leu
US-10-029-907-3

  Query Match             95.0%;  Score 1951;  DB 13;  Length 2201;
  Best Local Similarity   93.1%;  Pred. No. 4.4e-186;
  Matches  366;  Conservative   14;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| ||||||:||: |||||||||| ||| ||||||||||||||||||:
Db          5 MAASCGGAVFVGLILLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAV 64

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||:|||||| |||:|||| |||||||||||||||||| |||||||||| |||||
Db         65 ILLTCAIHPELIFTITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHY 124

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| |||||||||||||||:|||||||||||||||:||||| |:||||||||||
Db        125 VQMALMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAAC 184

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||| |||||: |||||||||||||||||||||||||||||||:|
Db        185 GDIILGLPVSARRGREIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRN 244

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db        245 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 304

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 364

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
Db        365 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 397


RESULT 8
US-10-309-561-3
; Sequence 3, Application US/10309561
; Publication No. US20030148348A1
; GENERAL INFORMATION:
;  APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
;  TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
;  TITLE OF INVENTION:  HEPATITIS C VIRUS
;  FILE REFERENCE: 13/083
;  CURRENT APPLICATION NUMBER: US/10/309,561
;  CURRENT FILING DATE:  2002-12-04
;  PRIOR APPLICATION NUMBER: US/10/029,907
;  PRIOR FILING DATE: 2001-12-21
;  PRIOR APPLICATION NUMBER: 60/257,857
;  PRIOR FILING DATE: 2000-12-22
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 2201
;   TYPE: PRT
;   ORGANISM: HCV
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: 882
;   OTHER INFORMATION: Xaa is Lys or Arg
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: 1489
;   OTHER INFORMATION: Xaa is Leu
US-10-309-561-3

  Query Match             95.0%;  Score 1951;  DB 14;  Length 2201;
  Best Local Similarity   93.1%;  Pred. No. 4.4e-186;
  Matches  366;  Conservative   14;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| ||||||:||: |||||||||| ||| ||||||||||||||||||:
Db          5 MAASCGGAVFVGLILLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAV 64

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||:|||||| |||:|||| |||||||||||||||||| |||||||||| |||||
Db         65 ILLTCAIHPELIFTITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHY 124

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| |||||||||||||||:|||||||||||||||:||||| |:||||||||||
Db        125 VQMALMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAAC 184

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||| |||||: |||||||||||||||||||||||||||||||:|
Db        185 GDIILGLPVSARRGREIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRN 244

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db        245 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 304

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 364

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
Db        365 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 397


RESULT 9
US-10-467-000-1
; Sequence 1, Application US/10467000
; Publication No. US20040067486A1
; GENERAL INFORMATION:
;  APPLICANT: De Francesco, Raffaele
;  APPLICANT:  Migliaccio, Giovanni
;  APPLICANT:  Paonessa, Giacomo
;  TITLE OF INVENTION: HEPATITIS C VIRUS REPLICONS AND REPLICON
;  TITLE OF INVENTION:  ENHANCED CELLS
;  FILE REFERENCE: ITR0003P
;  CURRENT APPLICATION NUMBER: US/10/467,000
;  CURRENT FILING DATE:  2003-07-21
;  PRIOR APPLICATION NUMBER: PCT/EP02/00526
;  PRIOR FILING DATE: 2002-01-16
;  PRIOR APPLICATION NUMBER: 60/263,479
;  PRIOR FILING DATE: 2001-01-23
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 3010
;   TYPE: PRT
;   ORGANISM: Con 1 HCV isolate nucleic acid
US-10-467-000-1

  Query Match             95.0%;  Score 1951;  DB 12;  Length 3010;
  Best Local Similarity   93.1%;  Pred. No. 6.9e-186;
```

Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

```
Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||||:|| ||||||:||: |||||||||| ||| ||||||||||||||||||:
Db        814 MAASCGGAVFVGLILLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAV 873

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||||||:|||||| |||:|||| |||||||||||||||||| |||||||||| |||||
Db        874 ILLTCAIHPELIFTITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHY 933

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||| |||||||||||||||:|||||||||||||||:||||| |:||||||||||
Db        934 VQMALMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAAC 993

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||| |||||: |||||||||||||||||||||||||||||||:|
Db        994 GDIILGLPVSARRGREIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRN 1053

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db       1054 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 1113

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1114 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 1173

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
Db       1174 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 1206
```

RESULT 10
US-09-919-901-4
; Sequence 4, Application US/09919901
; Publication No. US20030082518A1
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: :
US-09-919-901-4

Query Match 94.8%; Score 1946; DB 10; Length 1692;
Best Local Similarity 92.9%; Pred. No. 9.7e-186;
Matches 365; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

```
Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||||:|| |||||||||| |||||||||| || |||| ||||||| |||||||
Db         93 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFTTRAEAHLHVWIPPLNARGGRDAI 152

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| |||||||||||||:|| ||||||||||:|||||||||| |||||||| |||||
Db        153 ILLMCAVHPELIFDITKLLIAILGPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHY 212

Qy        121 VQMAFMKLAALTGTYVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 180
              |||||||| |||||:|:|||||:|||||||||||||||:|||| ||||||||||
Db        213 VQMAFMKLGALTGTYIYNHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 272

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||:|||||||: ||:||||||||||||||||||||||||||||||
Db        273 GDIILGLPVSARRGKEILLGPADSLEGRGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 332

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db        333 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 392

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              |||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        393 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 452

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||||:||||||||||
Db        453 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 485
```

RESULT 11
US-10-191-966-4
; Sequence 4, Application US/10191966
; Publication No. US20030175692A1
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/10/191,966
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: US/09/263,933
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: :
US-10-191-966-4

Query Match 94.8%; Score 1946; DB 14; Length 1692;
Best Local Similarity 92.9%; Pred. No. 9.7e-186;
Matches 365; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

```
Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||||:|| |||||||||| |||||||||| || |||| ||||||| |||||||
Db         93 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFTTRAEAHLHVWIPPLNARGGRDAI 152

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| |||||||||||||:|| ||||||||||:|||||||||| |||||||| |||||
Db        153 ILLMCAVHPELIFDITKLLIAILGPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHY 212

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||| |||||:|:|||||:|||||||||||||||:|||| ||||||||||
Db        213 VQMAFMKLGALTGTYIYNHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 272

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||||:|||||||: ||:||||||||||||||||||||||||||||||
Db        273 GDIILGLPVSARRGKEILLGPADSLEGRGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 332

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db        333 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 392

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              |||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        393 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 452
```

```
Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              ||||||||||||||||||||||:||||||||||
Db        453 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 485


RESULT 12
US-09-919-901-2
; Sequence 2, Application US/09919901
; Publication No. US20030082518A1
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT:  Jackson, Roberta L.
;  APPLICANT:  Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION:  OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/919,901
;  CURRENT FILING DATE:  2001-08-02
;  PRIOR APPLICATION NUMBER: 09/263,933
;  PRIOR FILING DATE: 1999-02-08
;  PRIOR APPLICATION NUMBER: 09/129,611
;  PRIOR FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 2307
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: :
US-09-919-901-2

  Query Match             94.8%;  Score 1946;  DB 10;  Length 2307;
  Best Local Similarity   92.9%;  Pred. No. 1.5e-185;
  Matches  365;  Conservative   13;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| |||||||||| |||||||||  || |||| ||||||| |||||||
Db        185 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFTTRAEAHLHVWIPPLNARGGRDAI 244

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| |||||||||||||:|| ||||||||||:|||||||||| ||||||| |||||
Db        245 ILLMCAVHPELIFDITKLLIAILGPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHY 304

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||| ||||||:|:|||||:|||||||||||||||||:||||| ||||||||||||
Db        305 VQMAFMKLGALTGTYIYNHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 364

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||:|||||||: ||:||||||||||||||||||||||||||||||||
Db        365 GDIILGLPVSARRGKEILLGPADSLEGRGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 424

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db        425 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 484

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 544

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              ||||||||||||||||||||||:||||||||||
Db        545 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 577


RESULT 13
US-10-191-966-2
; Sequence 2, Application US/10191966
; Publication No. US20030175692A1
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT:  Jackson, Roberta L.
;  APPLICANT:  Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION:  OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/10/191,966
;  CURRENT FILING DATE:  2002-07-10
;  PRIOR APPLICATION NUMBER: US/09/263,933
;  PRIOR FILING DATE: 1999-03-08
;  PRIOR APPLICATION NUMBER: 09/129,611
;  PRIOR FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 2307
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: :
US-10-191-966-2

  Query Match             94.8%;  Score 1946;  DB 14;  Length 2307;
  Best Local Similarity   92.9%;  Pred. No. 1.5e-185;
  Matches  365;  Conservative   13;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              ||||||||||:|| |||||||||| |||||||||  || |||| ||||||| |||||||
Db        185 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFTTRAEAHLHVWIPPLNARGGRDAI 244

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| |||||||||||||:|| ||||||||||:|||||||||| ||||||| |||||
Db        245 ILLMCAVHPELIFDITKLLIAILGPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHY 304

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||| ||||||:|:|||||:|||||||||||||||||:||||| ||||||||||||
Db        305 VQMAFMKLGALTGTYIYNHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 364

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||:|||||||: ||:||||||||||||||||||||||||||||||||
Db        365 GDIILGLPVSARRGKEILLGPADSLEGRGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 424

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db        425 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 484

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 544

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              ||||||||||||||||||||||:||||||||||
Db        545 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 577


RESULT 14
US-09-919-901-11
; Sequence 11, Application US/09919901
; Publication No. US20030082518A1
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT:  Jackson, Roberta L.
;  APPLICANT:  Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION:  OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/09/919,901
;  CURRENT FILING DATE:  2001-08-02
;  PRIOR APPLICATION NUMBER: 09/263,933
;  PRIOR FILING DATE: 1999-02-08
;  PRIOR APPLICATION NUMBER: 09/129,611
;  PRIOR FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 1692
```

```
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: :
US-09-919-901-11

  Query Match             94.6%;  Score 1943;  DB 10;  Length 1692;
  Best Local Similarity   92.6%;  Pred. No. 1.9e-185;
  Matches  364;  Conservative   14;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||||:|| |||||||||| |||||||||| || |||| ||||||| |||||||
Db         93 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFTTRAEAHLHVWIPPLNARGGRDAI 152

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| |||||||||||||||:|| ||||||||||:|||||||||| ||||||| |||||
Db        153 ILLMCAVHPELIFDITKLLIAILGPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHY 212

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||| |||||||:|:|||||:|||||||||||||||:||||| ||||||||||||
Db        213 VQMAFMKLGALTGTYIYNHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 272

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||:||||||||: ||:|||||||||||||||||||||||||||||||
Db        273 GDIILGLPVSARRGKEILLGPADSLEGRGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 332

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||
Db        333 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 392

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:|||||||||||||||||||||||||||||||||||||||||||:||||||||
Db        393 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSAGGPLLCPS 452

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
Db        453 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 485


RESULT 15
US-10-191-966-11
; Sequence 11, Application US/10191966
; Publication No. US20030175692A1
; GENERAL INFORMATION:
;  APPLICANT: Potts, Karen E.
;  APPLICANT:  Jackson, Roberta L.
;  APPLICANT:  Patick, Amy K.
;  TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
;  TITLE OF INVENTION:  OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
;  FILE REFERENCE: 0125-0005A
;  CURRENT APPLICATION NUMBER: US/10/191,966
;  CURRENT FILING DATE:  2002-07-10
;  PRIOR APPLICATION NUMBER: US/09/263,933
;  PRIOR FILING DATE: 1999-03-08
;  PRIOR APPLICATION NUMBER: 09/129,611
;  PRIOR FILING DATE: 1998-08-05
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 1692
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: :
US-10-191-966-11

  Query Match             94.6%;  Score 1943;  DB 14;  Length 1692;
  Best Local Similarity   92.6%;  Pred. No. 1.9e-185;
  Matches  364;  Conservative   14;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 MAASCGGAVFIGLALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
              |||||||||||:|| |||||||||| |||||||||| || |||| ||||||| |||||||
Db         93 MAASCGGAVFVGLVLLTLSPYYKVFLARLIWWLQYFTTRAEAHLHVWIPPLNARGGRDAI 152

Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| |||||||||||||||:|| ||||||||||:|||||||||| ||||||| |||||
Db        153 ILLMCAVHPELIFDITKLLIAILGPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHY 212

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||| |||||||:|:|||||:|||||||||||||||:||||| ||||||||||||
Db        213 VQMAFMKLGALTGTYIYNHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 272

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||:||||||||: ||:|||||||||||||||||||||||||||||||
Db        273 GDIILGLPVSARRGKEILLGPADSLEGRGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 332

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||
Db        333 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 392

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:|||||||||||||||||||||||||||||||||||||||||||:||||||||
Db        393 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSAGGPLLCPS 452

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
Db        453 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 485
```

```
Qy         61 ILLTCAVHPELIFDITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHY 120
              ||| |||||||||||||||:|| ||||||||||:|||||||||| ||||||| |||||
Db        153 ILLMCAVHPELIFDITKLLIAILGPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHY 212

Qy        121 VQMAFMKLAALTGTYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAC 180
              |||||||| |||||||:|:|||||:|||||||||||||||:||||| ||||||||||||
Db        213 VQMAFMKLGALTGTYIYNHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAAC 272

Qy        181 GDIISGLPVSARRGREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 240
              |||| |||||||||:||||||||: ||:|||||||||||||||||||||||||||||||
Db        273 GDIILGLPVSARRGKEILLGPADSLEGRGWRLLAPITAYSQQTRGLLGCIITSLTGRDKN 332

Qy        241 QVEGEVQVVSTATQSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 300
              ||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||
Db        333 QVEGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQA 392

Qy        301 PPGARSMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPS 360
              ||||||:|||||||||||||||||||||||||||||||||||||||||||:||||||||
Db        393 PPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSAGGPLLCPS 452

Qy        361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
              |||||||||||||||||||||:||||||||||
Db        453 GHAVGIFRAAVCTRGVAKAVDFVPVESMETTMR 485


Search completed: May  6, 2004, 09:43:19
Job time : 41.2025 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:08:45 ; Search time 51.7244 Seconds
(without alignments)
2075.771 Million cell updates/sec

Title: US-10-650-585-12
Perfect score: 1987
Sequence: 1 ALLTLSPYYKVLLARLIWWL.....RGVAKAVDFIPVESMETTMR 380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1987	100.0	380	5	ABG32185	Abg32185 HCV prote
2	1987	100.0	393	5	ABG32184	Abg32184 HCV prote
3	1987	100.0	409	5	ABG32181	Abg32181 HCV prote
4	1902	95.7	3010	2	AAR82694	Aar82694 Partial H
5	1897	95.5	3010	2	AAR68622	Aar68622 HCV prote
6	1896	95.4	3010	2	AAR68864	Aar68864 Hepatitis
7	1888	95.0	768	2	AAR40223	Aar40223 Recombina
8	1887	95.0	2201	5	ABG30601	Abg30601 Hepatitis
9	1887	95.0	2201	5	ABG30591	Abg30591 Hepatitis
10	1887	95.0	2201	5	ABG30600	Abg30600 Hepatitis
11	1887	95.0	2201	5	ABG30581	Abg30581 Hepatitis
12	1887	95.0	2201	5	ABG30593	Abg30593 Hepatitis
13	1887	95.0	2201	5	ABG30582	Abg30582 Hepatitis
14	1887	95.0	2201	5	ABG30580	Abg30580 Hepatitis
15	1887	95.0	2201	5	ABG30587	Abg30587 Hepatitis
16	1887	95.0	2201	5	ABG30599	Abg30599 Hepatitis
17	1887	95.0	2201	5	ABG30594	Abg30594 Hepatitis
18	1887	95.0	2201	5	ABG30598	Abg30598 Hepatitis
19	1887	95.0	2201	5	ABG30595	Abg30595 Hepatitis
20	1887	95.0	3010	5	ABG32458	Abg32458 Hepatitis
21	1887	95.0	3010	5	ABG32459	Abg32459 Hepatitis
22	1887	95.0	3010	5	ABG32451	Abg32451 Hepatitis
23	1887	95.0	3010	5	ABG32455	Abg32455 Hepatitis
24	1887	95.0	3010	5	ABG32457	Abg32457 Hepatitis
25	1887	95.0	3010	5	ABG32460	Abg32460 Hepatitis
26	1887	95.0	3010	5	ABG32461	Abg32461 Hepatitis
27	1887	95.0	3010	5	ABG32454	Abg32454 Hepatitis
28	1887	95.0	3011	5	ABG32456	Abg32456 Hepatitis
29	1884	94.8	2201	5	ABG30586	Abg30586 Hepatitis
30	1884	94.8	2201	5	ABG30589	Abg30589 Hepatitis
31	1884	94.8	2201	5	ABG30583	Abg30583 Hepatitis
32	1884	94.8	2201	5	ABG30588	Abg30588 Hepatitis
33	1883	94.8	2201	5	ABG30590	Abg30590 Hepatitis
34	1881	94.7	2307	3	AAY70064	Aay70064 Recombina
35	1881	94.7	3010	5	ABG32452	Abg32452 Hepatitis
36	1880	94.6	2201	5	ABG30584	Abg30584 Hepatitis
37	1880	94.6	2201	5	ABG30602	Abg30602 Hepatitis
38	1880	94.6	3010	5	ABG32453	Abg32453 Hepatitis
39	1878	94.5	2307	3	AAY70065	Aay70065 Recombina
40	1878	94.5	3014	2	AAR54099	Aar54099 NANBHV E1
41	1876	94.4	2201	5	ABG30585	Abg30585 Hepatitis
42	1873	94.3	3014	2	AAR35207	Aar35207 Hepatitis
43	1872	94.2	3090	7	ADD67962	Add67962 EMCV inte
44	1869	94.1	2307	3	AAY70066	Aay70066 Recombina
45	1869	94.1	3010	2	AAW98022	Aaw98022 Infectiou

ALIGNMENTS

RESULT 1
ABG32185
ID ABG32185 standard; protein; 380 AA.
XX
AC ABG32185;
XX
DT 05-NOV-2002 (first entry)
XX
DE HCV protease NS2/3 truncation mutant 827-1206.
XX
KW HCV; enzyme; protease; NS2/3 (827-1206); hepatitis C virus infection;
KW chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW chaotropic agent; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO200248375-A2.
XX
PD 20-JUN-2002.
XX
PF 13-DEC-2001; 2001WO-CA001796.
XX
PR 15-DEC-2000; 2000US-0256031P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
XX
DR WPI; 2002-599511/64.
XX
PT Novel polypeptide for screening inhibitors of non-structural proteases
PT useful as therapeutic agents against hepatitis C virus, comprises full
PT length non-structural protease, or its truncation.
XX
PS Claim 41; Page 60-61; 67pp; English.
XX
CC The invention relates to an isolated polypeptide consisting of a full-
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC residue amino acid 810 to 906, or having a minimal amino acid sequence
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC NS2/3 protease. Also included are (1) a composition (C) comprising an
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC its truncation or a mutated sequence, where the protease is in a solution
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide

CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC protease, involving isolating the protease in the presence of a
CC chaotropic agent, refolding the isolated protease by contacting it with a
CC reducing agent, and LDAO in the presence of reduced concentration of the
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC protease, involving diluting refolded inactive NS2/3 protease in a medium
CC containing an activation detergent to induce auto-cleavage of the NS2/3
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC protease, involving incubating the active NS2/3 protease produced by M2
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC cleavage products or their fragments, and measuring the presence or
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC active NS2/3 protease, involving carrying out M3 in the presence of, or
CC absence of the potential inhibitor, comparing the amount of uncleaved
CC NS2/3 protease, cleavage products or their fragments. The protease is
CC useful for detailed biochemical characterisation of the enzymes and in
CC the development of in vitro assays for screening novel inhibitors of
CC NS2/3 protease which are useful as therapeutic agents against HCV
CC infection (which causes chronic liver disease, cirrhosis and end-stage
CC liver disease. M1 is useful for high level production of protease. The
CC present sequence represents the NS2/3 truncation mutant 827-1206
CC (numbered relative to the full length NS2/3 protein)
XX
SQ Sequence 380 AA;

Query Match 100.0%; Score 1987; DB 5; Length 380;
Best Local Similarity 100.0%; Pred. No. 1.5e-184;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIF 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIF 60

Qy         61 DITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTG 120

Qy        121 TYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 180

Qy        181 GREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTAT 240

Qy        241 QSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCG 300

Qy        301 SSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCT 360

Qy        361 RGVAKAVDFIPVESMETTMR 380
              ||||||||||||||||||||
Db        361 RGVAKAVDFIPVESMETTMR 380
```

RESULT 2
ABG32184
ID ABG32184 standard; protein; 393 AA.
XX
AC ABG32184;
XX
DT 05-NOV-2002 (first entry)
XX
DE HCV protease NS2/3 truncation mutant 815-1206.
XX
KW HCV; enzyme; protease; NS2/3 (815-1206); hepatitis C virus infection;
KW chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW chaotropic agent; mutant; mutein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO200248375-A2.
XX
PD 20-JUN-2002.
XX
PF 13-DEC-2001; 2001WO-CA001796.
XX
PR 15-DEC-2000; 2000US-0256031P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
XX
DR WPI; 2002-599511/64.
XX
PT Novel polypeptide for screening inhibitors of non-structural proteases
PT useful as therapeutic agents against hepatitis C virus, comprises full
PT length non-structural protease, or its truncation.
XX
PS Claim 41; Page 59-60; 67pp; English.
XX
CC The invention relates to an isolated polypeptide consisting of a full-
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC residue amino acid 810 to 906, or having a minimal amino acid sequence
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC NS2/3 protease. Also included are (1) a composition (C) comprising an
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC its truncation or a mutated sequence, where the protease is in a solution
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC protease, involving isolating the protease in the presence of a
CC chaotropic agent, refolding the isolated protease by contacting it with a
CC reducing agent, and LDAO in the presence of reduced concentration of the
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC protease, involving diluting refolded inactive NS2/3 protease in a medium
CC containing an activation detergent to induce auto-cleavage of the NS2/3
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC protease, involving incubating the active NS2/3 protease produced by M2
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC cleavage products or their fragments, and measuring the presence or
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC active NS2/3 protease, involving carrying out M3 in the presence of, or
CC absence of the potential inhibitor, comparing the amount of uncleaved
CC NS2/3 protease, cleavage products or their fragments. The protease is
CC useful for detailed biochemical characterisation of the enzymes and in
CC the development of in vitro assays for screening novel inhibitors of
CC NS2/3 protease which are useful as therapeutic agents against HCV
CC infection (which causes chronic liver disease, cirrhosis and end-stage
CC liver disease. M1 is useful for high level production of protease. The
CC present sequence represents the NS2/3 truncation mutant 815-1206
CC (numbered relative to the full length NS2/3 protein)
XX
SQ Sequence 393 AA;

Query Match 100.0%; Score 1987; DB 5; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.6e-184;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIF 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         14 ALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIF 73

Qy         61 DITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         74 DITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTG 133

Qy        121 TYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 180
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        134 TYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 193

Qy        181 GREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTAT 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        194 GREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTAT 253

Qy        241 QSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCG 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        254 QSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCG 313

Qy        301 SSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCT 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        314 SSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCT 373

Qy        361 RGVAKAVDFIPVESMETTMR 380
                 ||||||||||||||||||||
Db        374 RGVAKAVDFIPVESMETTMR 393


RESULT 3
ABG32181
ID   ABG32181 standard; protein; 409 AA.
XX
AC   ABG32181;
XX
DT   05-NOV-2002  (first entry)
XX
DE   HCV protease NS2/3 (810-1206).
XX
KW   HCV; enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;
KW   chronic liver disease; cirrhosis; end-stage liver disease; virucide;
KW   hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KW   chaotropic agent; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         398. .409
FT                   /note= "Streptavidin tag"
XX
PN   WO200248375-A2.
XX
PD   20-JUN-2002.
XX
PF   13-DEC-2001; 2001WO-CA001796.
XX
PR   15-DEC-2000; 2000US-0256031P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Thibeault D,  Lamarre D,  Maurice R,  Pilote L,  Pause A;
XX
DR   WPI; 2002-599511/64.
DR   N-PSDB; ABK90406.
XX
PT   Novel polypeptide for screening inhibitors of non-structural proteases
PT   useful as therapeutic agents against hepatitis C virus, comprises full
PT   length non-structural protease, or its truncation.
XX
PS   Claim 42; Fig 1B; 67pp; English.
XX
CC   The invention relates to an isolated polypeptide consisting of a full-
CC   length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC   to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC   residue amino acid 810 to 906, or having a minimal amino acid sequence
CC   from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC   NS2/3 protease. Also included are (1) a composition (C) comprising an
CC   isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC   its truncation or a mutated sequence, where the protease is in a solution
CC   comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC   to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC   appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC   protease, involving isolating the protease in the presence of a
CC   chaotropic agent, refolding the isolated protease by contacting it with a
CC   reducing agent, and LDAO in the presence of reduced concentration of the
CC   chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC   protease, involving diluting refolded inactive NS2/3 protease in a medium
CC   containing an activation detergent to induce auto-cleavage of the NS2/3
CC   protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC   protease, involving incubating the active NS2/3 protease produced by M2
CC   for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC   cleavage products or their fragments, and measuring the presence or
CC   absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC   ; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC   active NS2/3 protease, involving carrying out M3 in the presence of, or
CC   absence of the potential inhibitor, comparing the amount of uncleaved
CC   NS2/3 protease, cleavage products or their fragments. The protease is
CC   useful for detailed biochemical characterisation of the enzymes and in
CC   the development of in vitro assays for screening novel inhibitors of
CC   NS2/3 protease which are useful as therapeutic agents against HCV
CC   infection (which causes chronic liver disease, cirrhosis and end-stage
CC   liver disease. M1 is useful for high level production of protease. The
CC   present sequence represents the NS2/3 (810-1206) protein, which has a C-
CC   terminal streptavividin tag
XX
SQ   Sequence 409 AA;

  Query Match             100.0%;  Score 1987;  DB 5;  Length 409;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-184;
  Matches  380;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIF 60
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         18 ALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIF 77

Qy         61 DITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTG 120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 DITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTG 137

Qy        121 TYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 TYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 197

Qy        181 GREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTAT 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 GREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTAT 257

Qy        241 QSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCG 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 QSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCG 317

Qy        301 SSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCT 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 SSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCT 377

Qy        361 RGVAKAVDFIPVESMETTMR 380
                 ||||||||||||||||||||
Db        378 RGVAKAVDFIPVESMETTMR 397


RESULT 4
AAR82694
ID   AAR82694 standard; protein; 3010 AA.
XX
AC   AAR82694;
XX
DT   16-OCT-2003  (revised)
DT   14-NOV-1996  (first entry)
XX
DE   Partial HCV non-structural polyprotein.
XX
KW   proteinase; hepatitis C virus; screening; inhibitor; proteolytic;
KW   identification; cleavage.
XX
```

```
OS   Hepatitis C virus; Virus.
XX
FH   Key             Location/Qualifiers
FT   Protein         898. .1233
FT                   /note= "partial proteinase; see AAR82692"
FT   Protein         992. .1907
FT                   /note= "partial proteinase; see AAR82693"
XX
PN   JP07184648-A.
XX
PD   25-JUL-1995.
XX
PF   05-FEB-1993;   93JP-00018854.
XX
PR   07-FEB-1992;   92JP-00022657.
PR   18-SEP-1992;   92JP-00249240.
PR   04-DEC-1992;   92JP-00325303.
XX
PA   (KAEN/) KAENNO K.
PA   (SUMQ ) SUMITOMO METAL IND LTD.
PA   (SOYA-) SOYAKU GIJUTSU KENKYUSHO KK.
XX
DR   WPI; 1995-287962/38.
DR   N-PSDB; AAT03960.
XX
PT   An HCV proteinase active substance - which has activity as an anti-HCV
PT   agent and can be used to screen for proteinase inhibitors.
XX
PS   Disclosure; Page 39-48; 52pp; Japanese.
XX
CC   The present sequence is a partial Hepatitis C Virus (HCV) polyprotein
CC   from the non-structral region. Partial proteinase sequences (AAR82692-
CC   93) are contained within this sequence. The proteinases can be used as
CC   anti-HCV agents. They can also be used to screen cpds. for their ability
CC   to inhibit their proteolytic activity. In this way proteinase inhibitors
CC   can be identified. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 3010 AA;

  Query Match            95.7%;  Score 1902;  DB 2;  Length 3010;
  Best Local Similarity  94.2%;  Pred. No. 5.3e-175;
  Matches  357;  Conservative   12;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          2 LLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIFD 61
              |||||||||| |||||||||| ||| |||||||:|||||||||||||||||||||||||
Db        828 LLTLSPYYKVFLARLIWWLQYFITRAEAHLQVWVPPLNVRGGRDAIILLTCAVHPELIFD 887

Qy         62 ITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGT 121
              |||||||| ||||||||||:|||||||||||||||||| ||||||||||||||||||||
Db        888 ITKLLLAILGPLMVLQAGITRVPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGT 947

Qy        122 YVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRG 181
              |||||||||:|||||||||||||||:||||| |:|||||||||||||||||||||||||
Db        948 YVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKLITWGADTAACGDIISGLPVSARRG 1007

Qy        182 REILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQ 241
              :|||||||:|  ||||||||||||||||||||||||||||||:|||||:|||||
Db       1008 KEILLGPADSFGEQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVDGEVQVLSTATQ 1067

Qy        242 SFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGS 301
              |||||||||||||:|||||||||||||||||||||||||||||| ||||||||||||||
Db       1068 SFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWPAPPGARSMTPCTCGS 1127

Qy        302 SDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTR 361
              |||||||||||:|||||||||||||:|||||||||||||||| ||||||||||
Db       1128 SDLYLVTRHADVVPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPSGHVVGIFRAAVCTR 1187

Qy        362 GVAKAVDFIPVESMETTMR 380
              |||||||||||||||||||
Db       1188 GVAKAVDFIPVESMETTMR 1206


RESULT 5
AAR68622
ID   AAR68622 standard; protein; 3010 AA.
XX
AC   AAR68622;
XX
DT   16-OCT-2003  (revised)
DT   16-OCT-1995  (first entry)
XX
DE   HCV protein cleavable with new serine proteinase.
XX
KW   proteinase; serine; cleavage; hepatitis C virus; HCV.
XX
OS   Hepatitis C virus; Virus.
XX
FH   Key             Location/Qualifiers
FT   Cleavage-site   2419. .2420
FT                   /note= "Serine protease cleavage site"
XX
PN   JP06315377-A.
XX
PD   15-NOV-1994.
XX
PF   06-MAY-1993;   93JP-00105666.
XX
PR   06-MAY-1993;   93JP-00105666.
XX
PA   (KAEN/) KAENNO K.
PA   (SUMQ ) SUMITOMO METAL IND LTD.
PA   (SOYA-) SOYAKU GIJUTSU KENKYUSHO KK.
XX
DR   WPI; 1995-032330/05.
DR   N-PSDB; AAQ80498.
XX
PT   New HCV-originated proteinase active substance - used for site-specific
PT   cleavage by an intermolecular reaction and the purification thereof.
XX
PS   Disclosure; Page 10-19; 23pp; Japanese.
XX
CC   This protein from HCV (hepatitis C virus) (encoded by AAQ80498) is
CC   cleaved between amino acids 2419 and 2420, by a new serine protease,
CC   contg. the sequence of AAR68621. The proteinase is purified as a fused
CC   product with the dihydrofolate reductase protein by using a methotrexate
CC   column. It can be used for the development of an inhibitor for HCV
CC   proteinase. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 3010 AA;

  Query Match            95.5%;  Score 1897;  DB 2;  Length 3010;
  Best Local Similarity  93.9%;  Pred. No. 1.6e-174;
  Matches  356;  Conservative   12;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 LLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIFD 61
              |||||||||| |||||||||| ||| |||||||:|||||||||||||||||||||||||
Db        828 LLTLSPYYKVFLARLIWWLQYFITRAEAHLQVWVPPLNVRGGRDAIILLTCAVHPELIFD 887

Qy         62 ITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGT 121
              |||||| | |||||||||||:|||||||||||||||||| ||||||||||||||||||||
Db        888 ITKLLLPILGPLMVLQAGITRVPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGT 947

Qy        122 YVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRG 181
              |||||||||:|||||||||||||||:||||| |:|||||||||||||||||||||||||
Db        948 YVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKLITWGADTAACGDIISGLPVSARRG 1007

Qy        182 REILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQ 241
              :|||||||:|  ||||||||||||||||||||||||||||||:|||||:|||||
Db       1008 KEILLGPADSFGEQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVDGEVQVLSTATQ 1067

Qy        242 SFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGS 301
              |||||||||||||:|||||||||||||||||||||||||||||| ||||||||||||||
Db       1068 SFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWPAPPGARSMTPCTCGS 1127

Qy        302 SDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTR 361
              |||||||||||:|||||||||||||:|||||||||||||||| ||||||||||
Db       1128 SDLYLVTRHADVVPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPSGHVVGIFRAAVCTR 1187
```

```
Db       1128 SDLYLVTRHADVVPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPSGHVVGIFRAAVCTR 1187

Qy        362 GVAKAVDFIPVESMETTMR 380
              ||||||||||||||||||
Db       1188 GVAKAVDFIPVESMETTMR 1206


RESULT 6
AAR68864
ID   AAR68864 standard; protein; 3010 AA.
XX
AC   AAR68864;
XX
DT   06-DEC-1995  (first entry)
XX
DE   Hepatitis C virus RNA helicase.
XX
KW   Hepatitis C virus; HCV; non-A non-B; helicase gene; RNA helicase;
KW   baculovirus; recombinant production.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   Region          196. .198
FT                   /label= N-linked glycosylation site
FT   Region          209. .211
FT                   /label= N-linked glycosylation site
FT   Region          234. .236
FT                   /label= N-linked glycosylation site
FT   Region          250. .252
FT                   /label= N-linked glycosylation site
FT   Region          305. .307
FT                   /label= N-linked glycosylation site
FT   Region          325. .327
FT                   /label= N-linked glycosylation site
FT   Region          417. .419
FT                   /label= N-linked glycosylation site
FT   Region          423. .425
FT                   /label= N-linked glycosylation site
FT   Region          430. .432
FT                   /label= N-linked glycosylation site
FT   Region          448. .450
FT                   /label= N-linked glycosylation site
FT   Region          532. .534
FT                   /label= N-linked glycosylation site
FT   Region          556. .558
FT                   /label= N-linked glycosylation site
FT   Region          576. .578
FT                   /label= N-linked glycosylation site
FT   Region          623. .625
FT                   /label= N-linked glycosylation site
FT   Region          645. .647
FT                   /label= N-linked glycosylation site
FT   Region          1213. .1215
FT                   /label= N-linked glycosylation site
FT   Region          1255. .1257
FT                   /label= N-linked glycosylation site
FT   Region          2041. .2043
FT                   /label= N-linked glycosylation site
FT   Region          2077. .2079
FT                   /label= N-linked glycosylation site
FT   Region          2240. .2242
FT                   /label= N-linked glycosylation site
FT   Region          2788. .2790
FT                   /label= N-linked glycosylation site
XX
PN   JP06319583-A.
XX
PD   22-NOV-1994.
XX
PF   18-SEP-1992;   92JP-00249241.
XX
PR   18-SEP-1992;   92JP-00249241.
XX
PA   (SOYA-) SOYAKU GIJUTSU KENKYUSHO KK.
XX
DR   WPI; 1995-040330/06.
DR   N-PSDB; AAQ81559.
XX
PT   of hepatitis C virus helicase gene in baculovirus - useful for large
PT   scale prodn. of RNA helicase.
XX
PS   Claim 1; Fig 1-4; 9pp; Japanese.
XX
CC   AAQ81559 encodes AAR68864 hepatitis C virus (HCV) RNA helicase. The DNA
CC   was used in the construction of an expression vector, which was used to
CC   transform a baculovirus host. The transformed baculovirus could then be
CC   used for the recombinant prodn. of HCV RNA helicase
XX
SQ   Sequence 3010 AA;

  Query Match            95.4%;  Score 1896;  DB 2;  Length 3010;
  Best Local Similarity  93.9%;  Pred. No. 2e-174;
  Matches  356;  Conservative   12;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 LLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIFD 61
              |||||||||| |||||||||| ||| |||||||:|||||||||||||||||||||||||||
Db        828 LLTLSPYYKVFLARLIWWLQYFITRAEAHLQVWVPPLNVRGGRDAIILLTCAVHPELIFD 887

Qy         62 ITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGT 121
              |||||||| ||||||||||:||||||||||||||||| ||||||||||||||||||||||
Db        888 ITKLLLAILGPLMVLQAGITRVPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGT 947

Qy        122 YVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRG 181
              |||||||||:|||||||||||||||||||:||||| |:|||||||||||||||||||||||
Db        948 YVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKLITWGADTAACGDIISGLPVSARRG 1007

Qy        182 REILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQ 241
              :|||||||:|   ||||||||||||||||||||||||||||||||:|||||:|||||||
Db       1008 KEILLGPADSFGEQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVDGEVQVLSTATQ 1067

Qy        242 SFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGS 301
              |||||||||||||:|||||||||||| |||||||||||||| ||||||||||||||||||
Db       1068 SFLATCVNGVCWTVYHGAGSKTLAGPKAPITQMYTNVDQDLVGWPAPPGARSMTPCTCGS 1127

Qy        302 SDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTR 361
              |||||||||||:|||||||||||||||||:|||||||||||||||| ||||||||||||||
Db       1128 SDLYLVTRHADVVPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPSGHVVGIFRAAVCTR 1187

Qy        362 GVAKAVDFIPVESMETTMR 380
              ||||||||||||||||||
Db       1188 GVAKAVDFIPVESMETTMR 1206


RESULT 7
AAR40223
ID   AAR40223 standard; protein; 768 AA.
XX
AC   AAR40223;
XX
DT   21-FEB-1994  (first entry)
XX
DE   Recombinant hepatitis C virus genomic protein.
XX
KW   Hepatitis; HCV; virus; screening; antiviral drugs.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 10
FT                   /note= "UUA encodes Ile."
FT   Misc-difference 81. .82
FT                   /note= "Nucleotide sequence encodes another Gly"
FT   Duplication     528. .768
FT                   /note= "Duplication of 241 amino acids at start of
FT                   protein sequence."
```

```
FT   Misc-difference 537
FT                   /note= "UUA encodes Ile."
XX
PN   JP05192160-A.
XX
PD   03-AUG-1993.
XX
PF   20-JAN-1992;   92JP-00028833.
XX
PR   20-JAN-1992;   92JP-00028833.
XX
PA   (BANY ) BANYU PHARM CO LTD.
XX
DR   WPI; 1993-277474/35.
DR   N-PSDB; AAQ48215.
XX
PT   Hepatitis C virus genomic RNA, cDNA and polypeptide - used for screening
PT   hepatitis C virus-specific antivirus drugs.
XX
PS   Claim 9; Page 4-6; 14pp; Japanese.
XX
CC   The protein is useful in the screening of HCV-specific antiviral drugs.
CC   HCV cDNA was cloned from plasma. Plasmids pSR3241 and pSR2541 were
CC   prepared using the cDNA and plasmid pSR3241 was used to transform a COS-1
CC   cell
XX
SQ   Sequence 768 AA;

  Query Match             95.0%;  Score 1888;  DB 2;  Length 768;
  Best Local Similarity   93.2%;  Pred. No. 1.8e-174;
  Matches  354;  Conservative   13;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 ALLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIF 60
              |||||||||| |||||||||| ||: |||||||:|||||||||||||||| |||||||||
Db         88 ALLTLSPYYKVFLARLIWWLQYFITKAEAHLQVWVPPLNVRGGRDAIILLMCAVHPELIF 147

Qy         61 DITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTG 120
              |||||||:| |||||||| : :|||||||||||||||||||||||||||||:|||||||
Db        148 DITKLLLSILGPLMVLQASLIRVPYFVRAQGLIRACMLVRKAAGGHYVQMAFVKLAALTG 207

Qy        121 TYVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 180
              |||||||||||||| |||||||||||:|| || |:||||||||||||||||||||||||
Db        208 TYVYDHLTPLQDWAHVGLRDLAVAVEPVVFSAMETKVITWGADTAACGDIISGLPVSARR 267

Qy        181 GREILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTAT 240
              |:|||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        268 GKEILLGPADSFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTAT 327

Qy        241 QSFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCG 300
              |||||||||| |||||||||||||||||||||||||||||||||| ||||||:||||||
Db        328 QSFLATCVNGACWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWPAPPGARSLTPCTCG 387

Qy        301 SSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCT 360
              |||||||||||||||||:|||||||||:|||||||||||||||| ||||||||||
Db        388 SSDLYLVTRHADVIPVRRRGDTRGSLLSPRPISYLKGSSGGPLLCPSGHVVGIFRAAVCT 447

Qy        361 RGVAKAVDFIPVESMETTMR 380
              ||||||||||||||||||||
Db        448 RGVAKAVDFIPVESMETTMR 467


RESULT 8
ABG30601
ID   ABG30601 standard; protein; 2201 AA.
XX
AC   ABG30601;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #10.
XX
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 882
FT                   /label= Arg, Lys
FT   Misc-difference 2183
FT                   /note= "Wild type Met substituted by Thr"
XX
PN   WO200252015-A2.
XX
PD   04-JUL-2002.
XX
PF   20-DEC-2001; 2001WO-CA001843.
XX
PR   22-DEC-2000; 2000US-0257857P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Kukolj G,  Pause A;
XX
DR   WPI; 2002-575382/61.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
XX
PS   Claim 3; Page; 140pp; English.
XX
CC   The invention describes a self-replicating hepatitis C virus (HCV)
CC   polynucleotide molecule comprising a 5'-non translated region (NTR),
CC   where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC   region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC   replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC   potential inhibitors of HCV replication. The HCV RNA molecule is also
CC   useful for efficiently establishing cell culture replication. The self-
CC   replicating polynucleotide molecule contains a 5'-NTR, where G at
CC   position 1 is substituted for A, and therefore provides an alternative to
CC   existing systems comprising a self-replicating HCV RNA molecule that, in
CC   conjunction with mutations in the HCV non-structural region, such as the
CC   G(2042)C/R mutations, transduces and/or replicates with greater
CC   efficiency. This amino acid sequence represents a mutant of the hepatitis
CC   C virus replicon APGK12 and contains the viral protease NS2/3, protease
CC   complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
CC   This sequence does not appear in the specification but has been created
CC   from the wild type sequence shown in ABG30580 using information given in
CC   the claims of the invention
XX
SQ   Sequence 2201 AA;

  Query Match             95.0%;  Score 1887;  DB 5;  Length 2201;
  Best Local Similarity   93.4%;  Pred. No. 9.9e-174;
  Matches  354;  Conservative   13;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          2 LLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIFD 61
              ||||||:||: |||||||||| ||| |||||||||||||||||:||||||:||||||
Db         19 LLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAVILLTCAIHPELIFT 78

Qy         62 ITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGT 121
              |||:|||| |||||||||||||||||| |||||||||| ||||||||| |||||||||
Db         79 ITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGT 138

Qy        122 YVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRG 181
              |||||||||:|||||||||||||||||:||||| |:||||||||||||| |||||||||
Db        139 YVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRG 198

Qy        182 REILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQ 241
              ||| |||||: |||||||||||||||||||||||||||||:|||||||||||||||||
Db        199 REIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQ 258

Qy        242 SFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGS 301
              |||||||||||||:||||||||||||||||||||||||||||||||||||:|||||||
```

```
Db        259 SFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGS 318

Qy        302 SDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTR 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 SDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTR 378

Qy        362 GVAKAVDFIPVESMETTMR 380
              ||||||||:||||||||||
Db        379 GVAKAVDFVPVESMETTMR 397


RESULT 9
ABG30591
ID   ABG30591 standard; protein; 2201 AA.
XX
AC   ABG30591;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #3.
XX
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 751
FT                   /note= "Wild type Ser substituted by Gly"
FT   Misc-difference 882
FT                   /label= Arg, Lys
XX
PN   WO200252015-A2.
XX
PD   04-JUL-2002.
XX
PF   20-DEC-2001; 2001WO-CA001843.
XX
PR   22-DEC-2000; 2000US-0257857P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Kukolj G,  Pause A;
XX
DR   WPI; 2002-575382/61.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
XX
PS   Claim 3; Page; 140pp; English.
XX
CC   The invention describes a self-replicating hepatitis C virus (HCV)
CC   polynucleotide molecule comprising a 5'-non translated region (NTR),
CC   where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC   region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC   replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC   potential inhibitors of HCV replication. The HCV RNA molecule is also
CC   useful for efficiently establishing cell culture replication. The self-
CC   replicating polynucleotide molecule contains a 5'-NTR, where G at
CC   position 1 is substituted for A, and therefore provides an alternative to
CC   existing systems comprising a self-replicating HCV RNA molecule that, in
CC   conjunction with mutations in the HCV non-structural region, such as the
CC   G(2042)C/R mutations, transduces and/or replicates with greater
CC   efficiency. This amino acid sequence represents a mutant of the hepatitis
CC   C virus replicon APGK12 and contains the viral protease NS2/3, protease
CC   complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
CC   This sequence does not appear in the specification but has been created
CC   from the wild type sequence shown in ABG30580 using information given in
CC   the claims of the invention
XX
SQ   Sequence 2201 AA;


  Query Match             95.0%;  Score 1887;  DB 5;  Length 2201;
  Best Local Similarity   93.4%;  Pred. No. 9.9e-174;
  Matches  354;  Conservative   13;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          2 LLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIFD 61
              ||||||:||: |||||||||| ||| ||||||||||||||||||:||||||:||||||
Db         19 LLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAVILLTCAIHPELIFT 78

Qy         62 ITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGT 121
              |||:|||| |||||||||||||||||| |||||||||| ||||||||| |||||||||
Db         79 ITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGT 138

Qy        122 YVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRG 181
              |||||||||:|||||||||||||||:||||| |:|||||||||||| |||||||||
Db        139 YVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRG 198

Qy        182 REILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQ 241
              ||| |||||: ||||||||||||||||||||||||||||||:|||||||||||||||
Db        199 REIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQ 258

Qy        242 SFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGS 301
              |||||||||||||:||||||||||||||||||||||||||||||||||||:|||||||
Db        259 SFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGS 318

Qy        302 SDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTR 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 SDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTR 378

Qy        362 GVAKAVDFIPVESMETTMR 380
              ||||||||:||||||||||
Db        379 GVAKAVDFVPVESMETTMR 397


RESULT 10
ABG30600
ID   ABG30600 standard; protein; 2201 AA.
XX
AC   ABG30600;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #9.
XX
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 882
FT                   /label= Arg, Lys
FT   Misc-difference 1357
FT                   /note= "Wild type Pro substituted by Leu"
XX
PN   WO200252015-A2.
XX
PD   04-JUL-2002.
XX
PF   20-DEC-2001; 2001WO-CA001843.
XX
PR   22-DEC-2000; 2000US-0257857P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Kukolj G,  Pause A;
XX
DR   WPI; 2002-575382/61.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
```

XX
PS Claim 3; Page; 140pp; English.
XX
CC The invention describes a self-replicating hepatitis C virus (HCV)
CC polynucleotide molecule comprising a 5'-non translated region (NTR),
CC where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC potential inhibitors of HCV replication. The HCV RNA molecule is also
CC useful for efficiently establishing cell culture replication. The self-
CC replicating polynucleotide molecule contains a 5'-NTR, where G at
CC position 1 is substituted for A, and therefore provides an alternative to
CC existing systems comprising a self-replicating HCV RNA molecule that, in
CC conjunction with mutations in the HCV non-structural region, such as the
CC G(2042)C/R mutations, transduces and/or replicates with greater
CC efficiency. This amino acid sequence represents a mutant of the hepatitis
CC C virus replicon APGK12 and contains the viral protease NS2/3, protease
CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
CC This sequence does not appear in the specification but has been created
CC from the wild type sequence shown in ABG30580 using information given in
CC the claims of the invention
XX
SQ Sequence 2201 AA;

Query Match 95.0%; Score 1887; DB 5; Length 2201;
Best Local Similarity 93.4%; Pred. No. 9.9e-174;
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

```
Qy          2 LLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIFD 61
              ||||||:||: |||||||||| ||| ||||||||||||||||:||||||:||||||
Db         19 LLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAVILLTCAIHPELIFT 78

Qy         62 ITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGT 121
              |||:|||| |||||||||||||||||||| |||||||||| |||||||| ||||||||
Db         79 ITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGT 138

Qy        122 YVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRG 181
              |||||||||:|||||||||||||||||:||||| |:|||||||||||| ||||||||
Db        139 YVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRG 198

Qy        182 REILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQ 241
              ||| |||||: ||||||||||||||||||||||||||||||:|||||||||||||||
Db        199 REIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQ 258

Qy        242 SFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGS 301
              |||||||||||||:|||||||||||||||||||||||||||||||||:|||||||
Db        259 SFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGS 318

Qy        302 SDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTR 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 SDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTR 378

Qy        362 GVAKAVDFIPVESMETTMR 380
              ||||||||:||||||||||
Db        379 GVAKAVDFVPVESMETTMR 397
```

RESULT 11
ABG30581
ID ABG30581 standard; protein; 2201 AA.
XX
AC ABG30581;
XX
DT 21-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #1.
XX
KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW cell culture replication; NS2/3; NS3/4; NS3; NS5B.
XX
OS Hepatitis C virus.
XX
PN WO200252015-A2.
XX
PD 04-JUL-2002.
XX
PF 20-DEC-2001; 2001WO-CA001843.
XX
PR 22-DEC-2000; 2000US-0257857P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI Kukolj G, Pause A;
XX
DR WPI; 2002-575382/61.
DR N-PSDB; ABK88573.
XX
PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT possess enhanced transduction or replication efficiency, useful for
PT evaluating potential inhibitors of HCV replication.
XX
PS Disclosure; Page 49-58; 140pp; English.
XX
CC The invention describes a self-replicating hepatitis C virus (HCV)
CC polynucleotide molecule comprising a 5'-non translated region (NTR),
CC where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC potential inhibitors of HCV replication. The HCV RNA molecule is also
CC useful for efficiently establishing cell culture replication. The self-
CC replicating polynucleotide molecule contains a 5'-NTR, where G at
CC position 1 is substituted for A, and therefore provides an alternative to
CC existing systems comprising a self-replicating HCV RNA molecule that, in
CC conjunction with mutations in the HCV non-structural region, such as the
CC G(2042)C/R mutations, transduces and/or replicates with greater
CC efficiency. This amino acid sequence is encoded by the hepatitis C virus
CC replicon APGK12 and contains the viral protease NS2/3, protease complex
CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B
XX
SQ Sequence 2201 AA;

Query Match 95.0%; Score 1887; DB 5; Length 2201;
Best Local Similarity 93.4%; Pred. No. 9.9e-174;
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

```
Qy          2 LLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIFD 61
              ||||||:||: |||||||||| ||| ||||||||||||||||:||||||:||||||
Db         19 LLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAVILLTCAIHPELIFT 78

Qy         62 ITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGT 121
              |||:|||| |||||||||||||||||||| |||||||||| |||||||| ||||||||
Db         79 ITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGT 138

Qy        122 YVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRG 181
              |||||||||:|||||||||||||||||:||||| |:|||||||||||| ||||||||
Db        139 YVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRG 198

Qy        182 REILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQ 241
              ||| |||||: ||||||||||||||||||||||||||||||:|||||||||||||||
Db        199 REIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQ 258

Qy        242 SFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGS 301
              |||||||||||||:|||||||||||||||||||||||||||||||||:|||||||
Db        259 SFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGS 318

Qy        302 SDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTR 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 SDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTR 378

Qy        362 GVAKAVDFIPVESMETTMR 380
              ||||||||:||||||||||
Db        379 GVAKAVDFVPVESMETTMR 397
```

RESULT 12
ABG30593

```
ID   ABG30593 standard; protein; 2201 AA.
XX
AC   ABG30593;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #4.
XX
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 882
FT                   /label= Arg, Lys
FT   Misc-difference 892
FT                   /note= "Wild type Leu substituted by Phe"
XX
PN   WO200252015-A2.
XX
PD   04-JUL-2002.
XX
PF   20-DEC-2001; 2001WO-CA001843.
XX
PR   22-DEC-2000; 2000US-0257857P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Kukolj G,  Pause A;
XX
DR   WPI; 2002-575382/61.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
XX
PS   Claim 3; Page; 140pp; English.
XX
CC   The invention describes a self-replicating hepatitis C virus (HCV)
CC   polynucleotide molecule comprising a 5'-non translated region (NTR),
CC   where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC   region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC   replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC   potential inhibitors of HCV replication. The HCV RNA molecule is also
CC   useful for efficiently establishing cell culture replication. The self-
CC   replicating polynucleotide molecule contains a 5'-NTR, where G at
CC   position 1 is substituted for A, and therefore provides an alternative to
CC   existing systems comprising a self-replicating HCV RNA molecule that, in
CC   conjunction with mutations in the HCV non-structural region, such as the
CC   G(2042)C/R mutations, transduces and/or replicates with greater
CC   efficiency. This amino acid sequence represents a mutant of the hepatitis
CC   C virus replicon APGK12 and contains the viral protease NS2/3, protease
CC   complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
CC   This sequence does not appear in the specification but has been created
CC   from the wild type sequence shown in ABG30580 using information given in
CC   the claims of the invention
XX
SQ   Sequence 2201 AA;

  Query Match             95.0%;  Score 1887;  DB 5;  Length 2201;
  Best Local Similarity   93.4%;  Pred. No. 9.9e-174;
  Matches  354;  Conservative   13;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          2 LLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIFD 61
              ||||||:||: |||||||||| ||| ||||||||||||||||||:||||||:||||||
Db         19 LLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAVILLTCAIHPELIFT 78

Qy         62 ITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGT 121
              |||:|||| |||||||||||||||||||| |||||||||| ||||||||| |||||||||
Db         79 ITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGT 138

Qy        122 YVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRG 181
              |||||||||:|||||||||||||||||:||||| |:|||||||||||||| |||||||||
Db        139 YVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRG 198

Qy        182 REILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQ 241
              ||| |||||: |||||||||||||||||||||||||||||||||:|||||||||||||||
Db        199 REIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQ 258

Qy        242 SFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGS 301
              |||||||||||||:|||||||||||||||||||||||||||||||||||||||:|||||||
Db        259 SFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGS 318

Qy        302 SDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTR 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 SDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTR 378

Qy        362 GVAKAVDFIPVESMETTMR 380
              ||||||||:||||||||||
Db        379 GVAKAVDFVPVESMETTMR 397


RESULT 13
ABG30582
ID   ABG30582 standard; protein; 2201 AA.
XX
AC   ABG30582;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #2.
XX
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 882
FT                   /note= "Wild type Lys substituted by Lys or Arg"
FT   Misc-difference 1233
FT                   /note= "Wild type Gly substituted by Cys"
XX
PN   WO200252015-A2.
XX
PD   04-JUL-2002.
XX
PF   20-DEC-2001; 2001WO-CA001843.
XX
PR   22-DEC-2000; 2000US-0257857P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Kukolj G,  Pause A;
XX
DR   WPI; 2002-575382/61.
DR   N-PSDB; ABK88574.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
XX
PS   Disclosure; Page 59-69; 140pp; English.
XX
CC   The invention describes a self-replicating hepatitis C virus (HCV)
CC   polynucleotide molecule comprising a 5'-non translated region (NTR),
CC   where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC   region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC   replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC   potential inhibitors of HCV replication. The HCV RNA molecule is also
CC   useful for efficiently establishing cell culture replication. The self-
CC   replicating polynucleotide molecule contains a 5'-NTR, where G at
CC   position 1 is substituted for A, and therefore provides an alternative to
CC   existing systems comprising a self-replicating HCV RNA molecule that, in
```

```
CC   conjunction with mutations in the HCV non-structural region, such as the
CC   G(2042)C/R mutations, transduces and/or replicates with greater
CC   efficiency. This amino acid sequence is encoded by the hepatitis C virus
CC   replicon APGK12 and contains the viral protease NS2/3, protease complex
CC   NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this
CC   sequence has been created from replicon APGK12 shown in ABG30581
XX
SQ   Sequence 2201 AA;

  Query Match             95.0%;  Score 1887;  DB 5;  Length 2201;
  Best Local Similarity   93.4%;  Pred. No. 9.9e-174;
  Matches  354;  Conservative   13;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          2 LLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIFD 61
              ||||||:||: |||||||||| ||| ||||||||||||||||||:|||||:||||||
Db         19 LLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAVILLTCAIHPELIFT 78

Qy         62 ITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGT 121
              |||:|||| |||||||||||||||||| |||||||||| ||||||||| |||||||||
Db         79 ITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGT 138

Qy        122 YVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRG 181
              |||||||||:|||||||||||||||||:||||| |:|||||||||||| |||||||||
Db        139 YVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRG 198

Qy        182 REILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQ 241
              ||| |||||: |||||||||||||||||||||||||||||||:||||||||||||||||
Db        199 REIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQ 258

Qy        242 SFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGS 301
              |||||||||||||:|||||||||||||||||||||||||||||||||||:|||||||
Db        259 SFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGS 318

Qy        302 SDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTR 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 SDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTR 378

Qy        362 GVAKAVDFIPVESMETTMR 380
              ||||||||:||||||||||
Db        379 GVAKAVDFVPVESMETTMR 397


RESULT 14
ABG30580
ID   ABG30580 standard; protein; 2201 AA.
XX
AC   ABG30580;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #9.
XX
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; NS2/3; NS3/4; NS3; NS5B.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 882
FT                   /note= "Encoded by ARG"
XX
PN   WO200252015-A2.
XX
PD   04-JUL-2002.
XX
PF   20-DEC-2001; 2001WO-CA001843.
XX
PR   22-DEC-2000; 2000US-0257857P.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.
XX
PI   Kukolj G,  Pause A;
XX
DR   WPI; 2002-575382/61.
XX
PT   New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT   possess enhanced transduction or replication efficiency, useful for
PT   evaluating potential inhibitors of HCV replication.
XX
PS   Disclosure; Page 69-74; 140pp; English.
XX
CC   The invention describes a self-replicating hepatitis C virus (HCV)
CC   polynucleotide molecule comprising a 5'-non translated region (NTR),
CC   where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC   region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC   replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC   potential inhibitors of HCV replication. The HCV RNA molecule is also
CC   useful for efficiently establishing cell culture replication. The self-
CC   replicating polynucleotide molecule contains a 5'-NTR, where G at
CC   position 1 is substituted for A, and therefore provides an alternative to
CC   existing systems comprising a self-replicating HCV RNA molecule that, in
CC   conjunction with mutations in the HCV non-structural region, such as the
CC   G(2042)C/R mutations, transduces and/or replicates with greater
CC   efficiency. This amino acid sequence is encoded by the hepatitis C virus
CC   replicon APGK12 and contains the viral protease NS2/3, protease complex
CC   NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B
XX
SQ   Sequence 2201 AA;

  Query Match             95.0%;  Score 1887;  DB 5;  Length 2201;
  Best Local Similarity   93.4%;  Pred. No. 9.9e-174;
  Matches  354;  Conservative   13;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          2 LLTLSPYYKVLLARLIWWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIFD 61
              ||||||:||: |||||||||| ||| ||||||||||||||||||:|||||:||||||
Db         19 LLTLSPHYKLFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAVILLTCAIHPELIFT 78

Qy         62 ITKLLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGGHYVQMAFMKLAALTGT 121
              |||:|||| |||||||||||||||||| |||||||||| ||||||||| |||||||||
Db         79 ITKILLAILGPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGT 138

Qy        122 YVYDHLTPLQDWAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRG 181
              |||||||||:|||||||||||||||||:||||| |:|||||||||||| |||||||||
Db        139 YVYDHLTPLRDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRG 198

Qy        182 REILLGPADNFEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQ 241
              ||| |||||: |||||||||||||||||||||||||||||||:||||||||||||||||
Db        199 REIHLGPADSLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQ 258

Qy        242 SFLATCVNGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSMTPCTCGS 301
              |||||||||||||:|||||||||||||||||||||||||||||||||||:|||||||
Db        259 SFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGS 318

Qy        302 SDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTR 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 SDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTR 378

Qy        362 GVAKAVDFIPVESMETTMR 380
              ||||||||:||||||||||
Db        379 GVAKAVDFVPVESMETTMR 397


RESULT 15
ABG30587
ID   ABG30587 standard; protein; 2201 AA.
XX
AC   ABG30587;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #7.
XX
KW   Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
KW   cell culture replication; NS2/3; NS3/4; NS3; NS5B.
XX
OS   Hepatitis C virus.
```